OPTIMAL OPERATION OF WATER RESOURCE SYSTEMS

A thesis presented for the
degree of Doctor of Philosophy in Civil Engineering
in the University of Canterbury,
Christchurch, New Zealand,

by

A.I. McKerchar

1971

ABSTRACT

The specification of a policy for the real-time operation of storages is viewed as part of the planning process for water resources development.

A policy-deriving algorithm, which can take account of complex stochastic hydrologies, uses the steps of streamflow simulation, deterministic dynamic programming, and linear regression to supply estimates of optimal policies.

Lag-one, log-normal Markov models, which are shown to fit monthly flow series from seven New Zealand rivers, are used to simulate synthetic data series. To include between-station correlations, a multistream model is developed.

The policy-deriving algorithm is applied to the year-by-year operation of an irrigation dam with the object of maximizing net returns, and to the month-by-month operation of two interconnected storages within a hydro-thermal power system with the object of minimizing thermal generating costs.

Derived linear policies, postulated to be near optimal for an uncertain future, caused expected operating costs for the hydro-thermal system which on average were 28% greater than the absolute minimum costs for a deterministic future. The cost difference, the "cost of uncertainty", could be reduced by using flow forecasts.

ACKNOWLEDGEMENTS

This study was undertaken within the Civil Engineering Department of Canterbury University; the support of this department and its professors has been greatly appreciated.

Professor John Burton of the Agricultural Engineering Department of Lincoln College (a constituent college of Canterbury University) acted as supervisor throughout the study. He provided inspiration and oversight, and the author is indebted to him.

The author must also thank the many staff members, both of Lincoln College and of Canterbury University, who have been of assistance at different times. In particular;

Mr Walter Boughton, who introduced the author to the subject matter and initiated the hydrologic analyses.

Mr F.P.S. Lu, who acted as joint supervisor in the latter stages of the work.

Dr David Painter, who read the thesis draft and made constructive and useful suggestions.

Mr Gerald Frengley, Mrs Mary Woods and Mr Rod Jensen, who provided useful comments at various stages.

The Computer Centre staff, both of Lincoln and Canterbury, who ran programs and gave assistance with programming difficulties.

Mrs O. Seebeck, who typed the thesis.

Thanks are also extended to Mr T. Lusk and Mr K. McCool of the New Zealand Electricity Department, and Mr G.G. Natusch and Mr H. Stewart of the Ministry of Works, all of whom helped in discussion and in providing information and data.

Financial assistance in the form of a postgraduate scholarship was provided by the University Grants Committee.

CONTENTS

PAGE

CHAPTER TWO

PAGE

LIST OF FIGURES

LIST OF TABLES

LIST OF COMMONLY OCCURRING SYMBOLS

SYMBOL	MEANING
A	(m x m) matrix
A_n	Surface area of storage in nth month
$a_o, a_1, a_2 \ldots.$	(auto) regression constants
B	(m x m) matrix
b_0, b_1, b_2	Regression constants
C_n	Thermal generation cost in nth month
C_s	Sample estimate of skewness for a stationary time series
C_{sj}	Sample estimate of skewness for jth months of a monthly series $(1 \leqslant j \leqslant 12)$
$c_0, c_1, c_2 \ldots$	Cost coefficients
D_A, D_B	Maximum monthly release from storage A, B.
D_M	Maximum release from a storage in a particular season.
D'	Maximum monthly discharge through power stations 3, 4, 5.
D''	Maximum monthly discharge through power stations 6, 7, 8.
d	Release from storage in a particular season.
d_A, d_B	Monthly release from storage A, B.
d_t	Release during tth season.
d'	Discharge through stations 3, 4, 5.
d''	Discharge through stations 6, 7, 8.
d_t^{antic}	Anticipated release in tth season.
d_t^{act}	Actual release in tth season.
E_n	Expected pan evaporation in nth month
e	% Absolute error in sample estimates of parameters.

SYMBOL	MEANING
$f_n(s)$	Present value of returns (costs) from optimally operating a storage through n time period, starting with a volume s in storage.
g_n	Quantity of thermal energy generated in nth month.
H_n	Total hydro-energy generated in nth month.
h_i	Hydro energy generated in ith station in nth month.
I_s	Shortage index.
j	Cyclic counter for months ($1 \leqslant j \leqslant 12$), j = 1 for January, 2 for February etc.
K_n	Energy load in nth month.
k	Time lag
L	Vector with components (x, y, z) describing spatial location of water, number of sites for possible development.
L*	Vector giving location of water demand.
L_n	Loss from storage in nth month.
L(s,d)	Loss from storage with initial vol. s, release d.
$L_A(s_A, s_A')$	Loss from storage A in nth month.
$L_B(s_B, s_B')$	Loss from storage B in nth month.
M	Number of possible water uses.
M_0	Matrix of lag-zero inter station correlations.
M_1	Matrix of lag-one inter station correlations.
m	Number of real stations, sample estimate of μ, number of discrete levels of flow.
N	Length of time series, number of possible irrigation districts.
n	Cumulative count of months, number of principal components.
P_i	Energy constant for ith power station.
p	Open water pan reduction factor.

SYMBOL	MEANING
p_j	Probability that inflow equals q_j.
$p_{ij}\ (q_i/q_j)$	Conditional probability that flow q_i follows flow q_j.
Q	Vector describing water quality, upper limit on total development.
Q^*	Vector giving possible water quality requirements.
$\bar{Q}$	Mean of q_t series.
$\bar{Q}_j$	Average flow in jth month.
$\bar{Q}_j^{\,p}$	Average flow in jth month at pth station.
$q,\ q_i,\ q_j$	Flow volumes.
q_n	Flow in nth month.
q_t	tth term of a time series of flow data.
$q_1,\ q_2,\ q_3,\ q_4$	Flows at stations 1, 2, 3, 4 respectively.
$q_t^{\,p}$	Flow in tth month at pth station.
q_L	Possible levels of development.
$\hat{q}_n$	Forecast of flow in nth month made at beginning of nth month.
R_n	Expected rainfall in nth month.
R_t	Return from releasing volume d_t in tth time period.
$R(d,s)$	Return from releasing volume d with storage initially containing volume s.
r	Interest rate.
$r(k)$	kth order serial correlation coefficient.
r_j	Serial correlation between flows in the jth and the previous month $(1 \leqslant j \leqslant 12)$.
S	Matrix representing properties of naturally occurring water, standard deviation of time series.
S^*	Matrix representing transformed state of availability.
$S_A,\ S_B$	Active storage volumes at A, B.

SYMBOL	MEANING
S_m	Active storage volume.
S_j	Standard deviation of flows in jth month.
S_j^p	Standard deviation of flows in jth month at pth station.
s	Beginning-of-period stored volume, sample estimate of standard deviation.
s'	End-of-period stored volume.
s_A, s_B	Beginning-of-month stored volumes at A, B.
s'_A, s'_B	End-of-month stored volumes at A, B.
s_t	Stored volume at beginning of tth period.
T	Vector giving availability of water in time.
T^*	Vector giving time distribution of water demands.
t_t	Pure random (0, 1) series.
$u_i\ (x_i)$	Net return from developing to level x_i at ith site.
$V(Q)$	Maximum net return from developing all sites up to a possible level Q.
$v_j\ (y_j)$	Net return from allocating a quantity y_j to jth use.
$w_k\ (z_k)$	Net return from allocating a quantity z_k to kth irrigation district.
X_n	Information known at beginning of nth month.
x_i	Storage volume developed at ith site.
$\hat{x}_i$	Optimal storage volume developed at ith site.
x_t	tth term of a time series.
$\bar{x}$	Average head at storage A in nth month.
$\bar{x}_Q$	Mean of estimates of synthetic means.
$\bar{x}_S$	Mean of estimates of synthetic standard deviations.
Y_t	Vector of standardized observations in tth time interval.
y_t^p	pth element of Y_t.
y_t	Stationary (0, 1) time series.

SYMBOL	MEANING
Z_t	Stationary time series remaining after removing persistence component from y_t.
z_α	Normal deviate corresponding to level of confidence α.
z_A, z_B	Spills at storages A, B.
z_k	Quantity allocated to kth irrigation district.
α	Level of confidence.
α_k	kth order autoregression coefficient.
γ_y	Population value of skewness for y_t series.
γ_z	Population value of skewness for Z_t series.
ϵ_t	tth term of pure random series, Vector of length m of random deviates.
Θ	Operator representing a system design.
θ	Frequency of second order correlogram.
μ	Population mean of log-transformed series.
$\rho(k)$	kth order autocorrelation coefficient.
$\rho^{(p)(q)}(k)$	lag-k cross correlation between stations p and q.
σ	Standard deviation of log-transformed series.
σ_Q, σ_S	Standard deviation of estimates of synthetic means, standard deviations.

CHAPTER ONE

A BASIS FOR PLANNING FOR WATER RESOURCES DEVELOPMENT

1.1 INTRODUCTION

In recent years planning for the development of water resources has become a matter of national and international importance. This is evidenced by the extensive and continuously expanding literature on the subject to be found in the professional journals of many disciplines.

Studies reported have drawn attention to the complex nature of water resource systems and to the need for a complete and systematic planning process which makes best use of limited water and capital resources.

This chapter describes the broad objectives of water resources development and outlines an idealized four step planning process. Subsequent discussion is centred on the third step of this process. This third step is described as a design step and involves answering questions about the sizes of structures to be built in the system, the levels of output to be achieved by the system, and the policy to be used in the operation of the system.

Before these questions can be answered, the objectives of the proposed development must be specified and translated into design criteria. If these criteria can be expressed quantitatively, optimizing techniques can be used to assist in the task of arriving at a best or optimal design.

The relative merits of available optimizing techniques are briefly discussed. A major limitation of all of these techniques is their inability to cope, either at a conceptual or at a computational level, or both, with stochastic (time-dependent probabilistic) hydrologic inputs.

One technique in particular, dynamic programming, is covered in some detail. Dynamic programming, a scheme for sequential decision making, appears to be the most suitable technique available for handling certain types of water resources design problems.

By coupling streamflow simulation with deterministic dynamic programming, Young (1966), explored a method of deriving operating policies while taking full account of the stochastic nature of the hydrologic inputs. Young did not show that the derived policies resembled the true optimum, but rather, that in some cases at least, the derived policy was better than classic operation policy rules. An example of a classic operation policy is described in Section 1.9.

This study parallels and extends the work of Young. In comparing the derived policy to the absolute optimum, the operating procedures for two real water resource systems are examined. These are:

(1) An irrigation dam and its year by year operation

(2) An interconnected hydro-thermal electric power system which incorporates two storage lakes and requires month-by-month operation decisions.

Chapter Two describes statistical models for representing monthly streamflow data. Variability in parameter estimates from short data records is examined empirically. A model which reproduces the basic statistics for the generation of a synthetic flow record is developed and verified for seven New Zealand rivers.

Chapter Three is concerned with the simulation of flow data for several flow stations within a region. A procedure which preserves the dependencies of records both backwards in time and between stations, as well as the statistical parameters of monthly flow volumes is developed and tested.

Chapter Four describes an application of dynamic programming and flow simulation to determine an operating policy for an irrigation dam.

In Chapter Five a model for deriving rules for the operation of a set of interconnected hydro-electric power stations is illustrated. Two large storage lakes are present and it is assumed that the system is to be operated in conjunction with a set of thermal stations to supply given monthly power loads while minimizing thermal generating costs. The deterministic algorithm is coupled with the multi-streamflow generating procedure to derive linear policy functions for the real-time operation of the lakes. Real-time operation with these policies is compared with the absolute optimum, which implies a full knowledge of future flow patterns.

A summary of the study and conclusions drawn from it are given in Chapter Six.

1.2 WATER RESOURCES DEVELOPMENT

Many water resource developments are large scale undertakings which involve big expenditures by the government of a country. The planning for such a development may be considered as a process of decision-making for public investment. Decisions must be made about which set of resources to develop, when to develop, the objectives to be met by the development and the particular water resource system design, or set of system designs, that will best meet the objectives.

A water resource system is defined herein as "an integrated group of physical structures and devices whose function is to convert naturally-occurring water and input resources of land, capital, labour and materials into controlled outputs of usable water at times, places and in quantities required for irrigation, water supply, power generation and other purposes." (Burton, 1964). The design of such a system involves the making of decisions as to the best sizes, locations and modes of operation of its components in order to achieve the desired objectives.

The complete design process involves an interplay of many disciplines, particularly economics and engineering. The questions "What can be done?" and "How much will it cost?" are best answered by the engineer; the questions "What is worth doing?" and "To what extent is it worthwhile?" by the economist, whilst it is the politician, acting on the answers of both engineers and economists, who must finally decide what is to be done, hopefully for the best advantage of the country.

The process of planning for water resources development can be set out as a four step process (Hufschmidt, 1965). An understanding of this process, which is summarized by Fig. 1.1, is useful in helping to identify the personnel, methodology, information and data needed to carry the process through to completion. The four steps are described individually.

1.2.1 Definition of Broad Objectives

The first step is to define that mix of broad objectives which is to be fulfilled by the development. This definition implies broad policy decisions by State or National Governments.

The broad objectives on which these policy decisions are based are fundamental goals relating to the welfare of society such as; increasing the net national income (economic efficiency), redistributing income, maintaining satisfactory levels of employment and economic growth, maintaining and enhancing cultural, aesthetic and historic values. In general, these objectives will conflict to some extent and it may not therefore be possible to design a water resource system which will fulfil each of a number of objectives to the highest possible degree. To maximize one objective, it may be necessary to sacrifice increments in other objectives.

A study of these objectives is beyond the scope of this thesis; detailed treatments are given by Eckstein (1958) and Maas et al (1962).

1.2.2 Translation of Objectives into Design Criteria

The second step of the planning process concerns the translation of the broad objectives of step one into quantitative design criteria. Hufschmidt (1965) gives the following as the more important criteria required for a typical design:

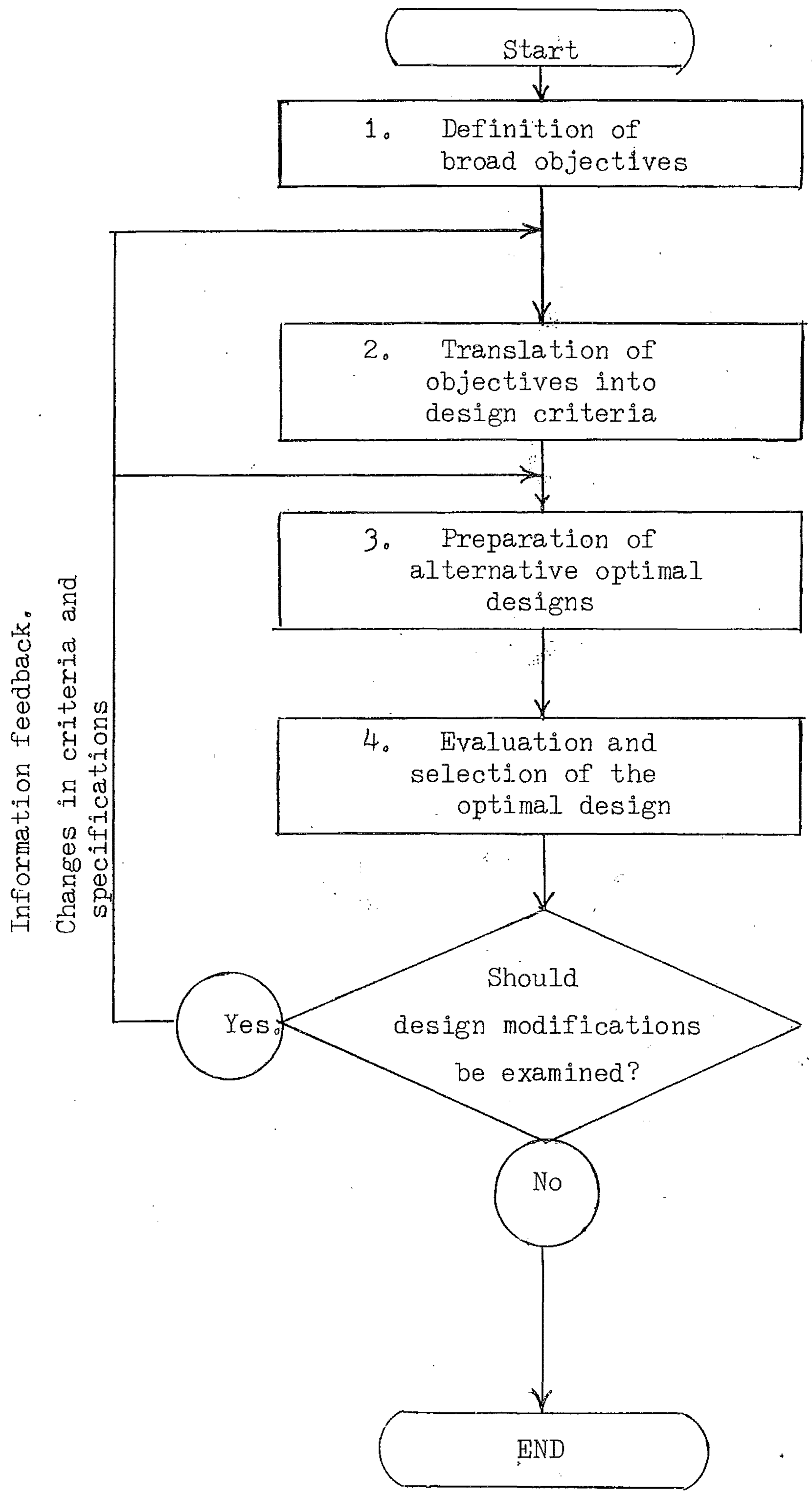
(a) The combination of objectives which are to be considered in the design, their relative weights and their form, either as values to be maximized or as constraints, making it possible to formulate an objective function.

(b) The discount rate.

(c) A weighting factor applied to capital and/or operation, maintenance and

Figure 1.1

ILLUSTRATION OF PLANNING PROCESS FOR WATER RESOURCES DEVELOPMENT



repair (O.M.R.) costs which will reflect the existence of budgeting constraints.

(d) A weighting factor to reflect the "opportunity cost of funds".

(e) A weighting factor which reflects divergencies between real labour costs and wages.

(f) Specific instructions for the construction of gross benefit functions for each major purpose of the water resource development.

(g) Specific instructions on how to handle uncertainty in hydrologic and other factors.

(h) Definition of the geographical region and of the scope of the design problem.

(i) Constraints concerned with water quality, political and institutional factors and other aspects.

1.2.3 Preparation of Alternative Optimal Designs

With the design criteria specified, step three involves the preparation of alternative system designs which optimize the value of the objective function, subject to the constraints and general rules set by the design criteria. At the planning stage a system design does not imply a detailed engineering design of system components, but rather a set of decisions about component sizes, levels of outputs, and policies for operating the system.

This chapter outlines a particular optimizing technique which can assist in preparing optimal designs. The aim of this thesis is to test and evaluate the technique.

1.2.4 Evaluation and Selection of Optimal Designs

The fourth step involves the evaluation of the alternative designs in the light of intangible non-quantifiable objectives which cannot be included in the optimizing process. These value judgements are a task for political policy-makers who at this stage become aware of the consequences of different designs and the sensitivity of the plants to variations in different criteria. As a result of this examination, changes may be made in the design criteria given in the second step and thus the design prepared in ~~step three~~ may need revision.

In this way a feedback of information from step four to steps two and three occurs; this allows the cycle of the design process, resource allocation and criteria specification to be completed.

It is the task of the political policy and decision-makers to decide the extent to which the designs satisfy the objectives set out in step one and to select the set which is best, in terms of these objectives. It may be that the development is not justified and that the objectives can be better met by other action. Again, this step, which is beyond the scope of this thesis, is covered in detail elsewhere. (McKean, 1958).

In practice this idealized process is barely approached. In both Australia and New Zealand, planning for water resources development has, in the past, tended to proceed in a piece-meal, rule-of-thumb manner. (Burton, 1964), (Jensen, 1968). There are however encouraging signs that the planning process is being improved.

This outline of the idealized planning process gives an introduction to the design problem described in step three. The aim of this thesis is to develop and test a method for tackling part of this step.

1.3 THE DESIGN STEP

Formally, the design step may be described as follows: (Buras and Herman, 1968).

The properties of naturally occurring water resources may be described by a matrix S consisting of 3 vectors,

$$S = \begin{bmatrix} L \\ T \\ Q \end{bmatrix},$$

where L is a vector with 3 components (x, y, z) which describe the spatial location of the water.

T is a vector giving the availability in time of the water. It consists of parameters of the probability distributions which describe the occurrence of the resource in time.

Q is a vector describing the quality of the water. It is partitioned into 3 sets of elements which respectively describe biological, chemical and physical properties. In general, each of these will be characterized by time - varying parameters of probability distributions.

Water resources development consists of transforming the original availability matrix S to another which fulfils the criteria listed in step two. Thus S becomes S*, which again consists of 3 vectors,

$$S^* = \begin{bmatrix} L^* \\ T^* \\ Q^* \end{bmatrix},$$

where L^* represents the location of the demand for water.

T^* represents the quantities of water and water derivatives demanded and the distribution of this demand in time.

Q^* represents the possible quality requirements.

This transformation of an available state into a desired state is achieved by an operator θ which is defined by the equation,

$$S^* = \theta . S.$$

The operator θ is in fact the design of the system transferring water from one state of availability to another. The design consists not only of physical structures for the collection, conveyance, storage, treatment and utilization of water, but also of the policy which states how much water to handle in each time period, the manner in which water is to be allocated to different uses and the way in which water derivatives will be allocated and used.

The design of the system is determined by the relationship between L, T and Q, and L*, T* and Q*. Differences in elevation between the water source and possible tail-water levels and differences between the water supply and power demand patterns will be factors in establishing the feasibility of hydroelectric power generation. Elevation differences between sources of water and demand points for water supply will establish whether gravity supply is suitable or whether pumping is required. The size of storage structures will be dependent on the differences between T and T*. The specifications for treatment facilities will be dictated by differences between T and T* and Q and Q*.

It is clear that inter-relationships will exist between these vectors; for example, the requirements of T*, which may call for a period of storage, could result in a change in the water quality Q.

The design problem, described by step three, is to choose θ such that S* is satisfied in an optimal manner.

Many examples exist in which it is difficult or impossible to express quantitatively the benefits of a project and to derive a numerical measure of the objective function which is to be maximized. For instance, the main aim of a project may be to contribute to public health and recreation. It is difficult to rank these objectives numerically by an objective function. The specified design criteria may necessarily be somewhat arbitrary, and in the design problem may be expressed in the form of constraints such as a minimum stream discharge or a maximum water level fluctuation to be tolerated, or an accepted level of risk.

However, there are also a number of types of projects where it is possible to imput monetary values to outputs such as irrigation water delivered and flood damages reduced. In such cases, net benefits can be evaluated quantitatively. Economic values can be attached to physical criteria and a comparison can be made between the increments in benefits resulting from a change in any criterion, and the incremental cost of achieving such a change.

With projects of this type, alternative designs can be ranked quantitatively according to the extent to which their objectives are fulfilled, but subject to, and constrained by, those parts of the design criteria not included in the ranking function. In this way the design process can select an optimum design; by finding that design, or set of designs, which best fulfil the objectives which initiated the process.

Although the benefits resulting from alternative designs may form a convex set, it may be that another convex set of benefits will exist when certain key parameters such as the interest rate are varied. While the possible values of these parameters should be specified in the design criteria at the highest decision making level, often they can only be properly arrived at if a number of designs having a range of values of these parameters are prepared. Furthermore, to allow for the variation in annual benefits due to the stochastic nature of the hydrology, some measure of risk, which will weight benefits from alternative designs should be introduced. This illustrates the close relationship between the four steps of the planning process and the importance of the information feedback described previously.

From an engineering design point of view, the design problem may be considered as comprising three closely connected decisions. (Buras, 1966).

(1) The optimal dimensions of structures to belong to the system.

(2) The optimal target output of the system, consistent with the objectives.

(3) The establishment of an operating policy for the system which will ensure that the objective function is optimized.

1.4 OPTIMIZING TECHNIQUES

Even when the design criteria can be quantitatively specified, and the design problem can be expressed mathematically as a problem of maximizing a numerical objective function subject to a given set of constraints, the task of finding an optimum design is still a demanding one.

Two basic approaches have emerged for tackling this optimization problem (Buras and Herman, 1968). In one, given an objective function, a set of constraints, and a mathematical formulation of the hydrology, an optimal solution is obtained directly by a mathematical algorithm. In the other, the behaviour of the system is simulated by routing a real or synthetic hydrologic record through a mathematical model of the proposed system and evaluating the performance of each element. These will be referred to as analytic and simulation approaches respectively. Examples of analytic solutions are given by Little (1955) and Buras (1965); applications of simulation are described by Morrice and Allan (1959), Lewis and Shoemaker (1962) and Hufschmidt and Fiering (1966).

1.4.1 Simulation Approach

With simulation, the description of the system, represented by mathematical model, may be complex and approach reality. At the same time a full statistical treatment of the hydrology of the system may be possible. In a simulation study, temporal sequences of events are studied within a computer where seconds can represent years and while in theory many variables may be optimized, programming and computational difficulties can easily arise. For example, a river basin in which several structures are planned may present say 20 variables (heights of dams, projected water deliveries and the like) from which the best combination is to be selected. If each variable can assume five different values, then in theory 5^{20} designs must be evaluated.

If systematic search techniques, such as the method of steepest ascent, are used to search the multi-dimensional response surface, which represents the objective function in such a case, no guarantee is given that the optimum arrived at is the global maximum and not a local maximum. Furthermore, the result does not offer a real insight as to the structure of the optimal solution.

In theory each combination of structures and outputs has associated with it an optimal operating policy. Though the simulation approach can be used to assist in making decisions about the sizes of structures and levels of output, it is inadequate for deriving optimal operating policies. Either a large part of the computer program used in the simulation may need to be changed for a small change in the operating policy, or a general program must be coded with parameters which may be varied to test different policies. In both cases the burden of computer programming can become excessive or impossible.

1.4.2 Analytic Approach

In contrast, analytic solutions use features of the probabilistic structure of the hydrology, as well as an objective function and its attendant constraints, to form an algorithm which can in simple cases be solved for any set of numerical values.

Within a very good review of the possible applications of mathematical programming in water resources engineering, Buras and Herman (1968) identify five basic analytic techniques; the calculus of variations, probability techniques such as queuing theory, linear programming, non-linear programming, and dynamic programming. Modifications of these techniques have given methods with special properties, such as dynamic linear programming and incremental dynamic programming, that are applicable to certain types of problems.

Applications of analytic techniques to water resource planning problems have been extensively reviewed by Buras (1966), Buras and Herman (1968), and Roefs (1968).

Some techniques, such as linear programming, are well suited to handling multi-dimensional problems, but cannot cope with complicated stochastic hydrologies or multi-period problems.

On the other hand, dynamic programming is well suited to multi-period optimizing problems, but is restricted by present day computers to optimizing only two or three variables. It is not restricted to linear or strictly concave objective functions. Instead, it can handle any type of function, which if tabulated, does not even need to be known in functional form. In certain forms the technique can cope with complex hydrologies which may show interdependencies both in space and time. In contrast to simulation and certain types of non-linear programming, one is assured that the optimum obtained by a dynamic programming algorithm is the global optimum for the problem in question.

It should be noted that simulation and analytic approaches are not mutually exclusive. Dorfman (1965) for example, suggests that initial linear programming solutions, obtained for an approximate model of a system, are useful in defining regions within which a detailed simulation study may be made.

1.5 THE DYNAMIC PROGRAMMING APPROACH

Given suitable simplifications and assumptions, most design problems can be formulated for solution by an analytic technique. Hall and Buras (1961) show how a general design problem, which involves answering the three levels of questions outlined previously, can be tackled by the dynamic programming approach.

The particular set of inter-related questions examined were "What plausible sites for reservoirs should be developed and to what extent?" "How should the storage space in each reservoir be allocated with respect to purpose. (e.g. irrigation, flood control)?" "How should water stored for irrigation be allocated to possible geographic areas?" Questions similar to the last could be asked for power, flood control, etc. and these must be answered to provide information for the second question. The first question needs answers to the second for all feasible reservoir sites.

This implies a vertical interdependence of levels of the problem and is described as a hierarchy of decisions. Once answers have been obtained at one level, the next higher level of decisions can be analysed.

At the highest decision level, the questions of which reservoir sites were to be developed and the extent of development were considered. Sites were numbered 1, 2, 3 ..., i, ... L and the function $u_i\ (x_i)$ denoted the net return from the ith site as a function of its storage volume x_i. The problem was to determine that set x_i which maximized the sum of net returns from all possible sites, subject to the restriction of an upper limit Q beyond which no water was available for storage, even in the long run.

Thus the maximum return for all sites, is

$$V(Q) = \max \left(\sum_{i=1}^{L} u_i\ (x_i) \right) \tag{1-1}$$

subject to

$$x_i \geqslant 0 \text{ and } \sum_{i=1}^{L} x_i \leqslant Q$$

Given deterministic inflows this problem can be tackled recursively by considering one, two, three ... sites remaining to be developed.

For one site remaining, define the vector,

$$f_1 (q_1) = \max (u_1 (x_1)), \tag{1-2}$$

subject to

$$0 \leqslant x_1 \leqslant q_1 \text{ for all } q_1, \ 0 \leqslant q_1 \leqslant Q,$$

as the maximum net return when a volume q_1 is available for allocation at the last site.

Here q_1 represents possible storage volumes, or levels of development at the last site, x_1 the amount undertaken if this site was the last remaining to be developed, and $u_1 (x_1)$ the net return from allocating an volume x_1. $f_1 (q_1)$ is defined for all feasible values of q_1. The value of x_1, which maximizes (1-2), say $\hat{x}_1$, represents an optimal allocation of the volume q_1 at the last site. $\hat{x}_1$ may or may not equal q_1. If it does not, a volume $(q_1 - \hat{x}_1)$ remains, which it is not economic to allocate.

Next, with two sites remaining for development, define the vector -

$$f_2 (q_2) = \max (u_2 (x_2) + f_1 (q_2 - x_2)), \tag{1-3}$$

subject to

$$0 \leqslant x_2 \leqslant q_2 \text{ for all } q_2, \ 0 \leqslant q_2 \leqslant Q,$$

as the net return from optimally allocating a volume q_2 over the last two sites. Here x_2 is the level of use at site 2, leaving an amount $q_1 = q_2 - x_2$ available at the last site, numbered 1. $\hat{x}_2$ is the optimal level of use, which maximizes (1-3). Allocation for the last site is defined by (1-2).

With Bellman's principle of optimality, (1-3) can be generalized for i sites remaining. Thus,

$$f_i (q_i) = \max (u_i (x_i) + f_{i-1} (q_i - x_i)), \tag{1-4}$$

subject to

$$0 \leqslant x_i \leqslant q_i \text{ for all } q_i, \ 0 \leqslant q_i \leqslant Q,$$

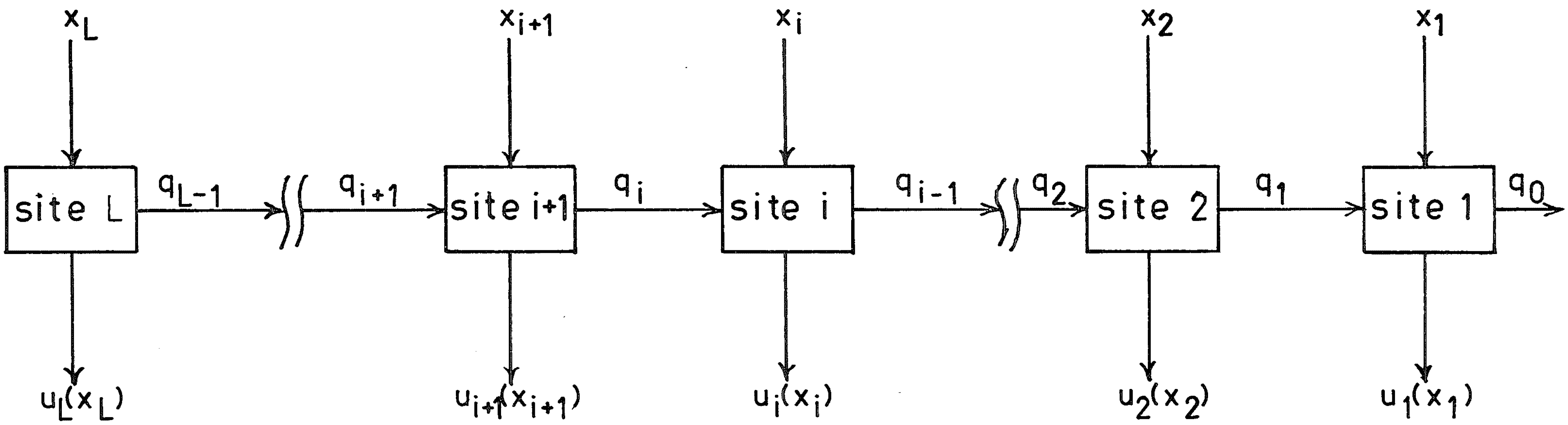
is the net return from optimally allocating q_i over i sites.

The process is shown graphically in Fig. 1.2. In this figure;

$u_i (x_i)$ is the return from allocating x_i at site i, and the objective is to maximize $\sum_{i=1}^{L} u_i (x_i)$.

x_i is a decision and is known as the decision variable.

FIG. 1.2 MULTI-STAGE DECISION MAKING



Sites represent stages, in this example there are L stages.

q_i is the state variable, defining the state of the system at the start of the ith stage.

At the final stage L, the vector $f_L(q_L)$ gives the maximum returns available by starting the L stage allocation process with a volume q_L. An optimal policy, or sequence of decisions can then be obtained.

Starting at the Lth stage with q_L, the optimal decision $\hat{x}_L$ results in $q_{L-1} = (q_L - \hat{x}_L)$ being available for the remaining (L-1) stages. By successively applying the relationship $q_{i-1} = (q_i - \hat{x}_i)$ and selecting the appropriate element from the x_{i-1} tables, a sequence of $\hat{x}_i$ values is obtained which defines the optimal allocation policy.

At the ith site, water will be allocated to M different uses so that the return from the storage of capacity x_i is maximized. If $v_j\ (y_j)$ is the net from allocating a volume of water y_j to the jth use,

$$u_i\ (x_i) = \max \left(\sum_{j=1}^{M} v_j\ (y_j) \right) \tag{1-5}$$

This can be solved by considering the allocation of water to successively 1, 2, 3 ...M remaining uses, and setting up and recursively solving an equation similar to (1-4). A solution to (1-5) is required for every possible level of storage at every possible site.

Finally, the sub-uses of water are considered. If, for instance, the jth use is irrigation, and $w_k\ (z_k)$ is the return from allocating a volume z_k to the kth irrigation district,

$$v_j\ (y_j) = \max \left(\sum_{k=1}^{N} w_k\ (z_k) \right) \tag{1-6}$$

The solution to (1-6) can also be obtained by setting up and solving another recursion equation. This solution is required for all uses and all storage volumes at all possible storage sites.

The setting up of this very general framework and the task of formulating a particular design problem within it may be described as the conceptual problem, and the equations derived, such as (1-1) to (1-6), may be termed the conceptual equations. The task of recursively solving these equations for a particular set of numerical values may be described as the computational problem.

Most of the progress in applying dynamic programming in water resources planning has been at the conceptual level. Solutions to computational problems have been limited by the capacity of present day (1970) computers to comparatively simple examples, involving at the most either one storage and several uses of water, or one use and two or three storages.

Further difficulties which hinder the direct application of the formulation described to particular problems are:

(1) The separability requirement that returns from each decision must be independent of other decisions.

(2) The complication caused by the uncertainty which surrounds future inflows.

Much attention has been given to preparing conceptual equations which cope with these difficulties for particular problems.

An example of how the first can be handled is given by Hall (1964). The problem was that of optimizing the design of a multiple purpose reservoir which was to provide releases for several purposes such as irrigation and hydro-electric power generation, as well as providing empty storage to reduce flood peaks.

Non-consumptive uses of water, such as hydro-power generation, can to some extent use the same water which may later be used in a consumptive use such as irrigation. This invalidates the separability requirement.

A numerical example of optimizing the operation of a multiple purpose reservoir is given by Hall et al (1968). A solution to the operation of a pumped storage scheme is given by Hall and Roefs (1966).

These studies were based on a critical period concept; the optimal operating policies were derived for historical sequences of runoff which were considered to be critical in the future operation of the system. In effect the policy derived indicated how the system should be operated given that a particular set of recorded flows would recur.

The use of critical period hydrology is one method of coping with the stochastic nature of the inflows and is of utility in the specification of firm levels of water supply and power production.

A major criticism is the implicit assumption that future inflow sequences during the real-time operation of the system will be represented by those which occurred historically. In fact a critical period may not necessarily be representative of future inflow sequences, especially if the historic record is brief, and the data show wide variability, as may well be the case in arid areas.

Two other approaches to handling the problem of uncertainty in hydrologic data are available. The first involves a direct stochastic dynamic programming solution, the second a simulation approach coupled with deterministic dynamic programming. These two approaches will be examined separately.

1.6 OPTIMAL POLICIES BY STOCHASTIC DYNAMIC PROGRAMMING

In its simplest form, the functional equation of dynamic programming was given by (1-4). When applied to reservoir operation, assuming storage sizes and output levels are fixed, and working backwards through time on an annual basis, it takes the following form:

$$f_n(s) = \max \left(R(d,s) + \frac{1}{1+r} f_{n-1}(s - d - L(s,d) + q) \right), \qquad (1\text{-}7)$$

$$0 \leqslant d \leqslant \min\left((s + q - L(s,d)), D_M \right)$$

$$0 \leqslant s \leqslant S_M$$

where r is the interest rate

$f_n(s)$ is the present value of returns from operating the storage of n future years, starting with a storage level s.

d is the release made during the nth year - the decision variable.

$R(d,s)$ is the return in period n from releasing a quantity d.

q is the inflow during the nth year.

$L(s,d)$ represents storage losses by seepage and evaporation.

$f_{n-1}(s-d-L(s,d)+q)$ is the maximum present value from operating the system for (n-1) future years.

S_M is the capacity of the storage.

D_M is the maximum release.

Let s' be the volume in storage at the end of the nth year,

thus $s' = s - d - L(s,d) + q$

$$(0 \leqslant s' \leqslant S_M)$$

s and s' are termed state variables, s is the initial state and s' the final state. The volume d released during the nth period is termed the decision variable. The nth period is referred to as the nth stage. Equation (1-7) can be rewritten as:

$$f_n(s) = \max (R(d,s) + \frac{1}{1+r} f_{n-1}(s')), \tag{1-8}$$

subject to the constraints of (1-7). The expression of the final state as a function of the initial state and the decision, that is

$$s' = s'(s,d),$$

is a fundamental requirement for the application of dynamic programming to any problem.

The formulation (1-8) assumes that future inflows are known in advance. In fact, future flows are not known with certainty. Unless flow data for some historic time such as a critical period are used, this formulation is of no immediate use.

Inflow is a stochastic variable which can be characterised by a probability distribution. If the approximation of a discrete distribution is used and $f_n(s)$ is redefined as the expected present value of returns from n years operation and p_j is defined as the probability that the flow will equal q_j, then equation (1-7) can be rewritten as:

$$f_n(s) = \max (R(d,s) + \frac{1}{1+r} \sum_{j=1}^{m} p_j \cdot f_{n-1}(s-d-L(s,d) + q_j)), \tag{1-9}$$

subject to

$$0 \leqslant d \leqslant \min((s + q_j - L(s,d)), D_M)$$

$$0 \leqslant s \leqslant S_M$$

$$\sum_{j=1}^{m} p_j = 1$$

This type of formulation, but with three state variables and three decision variables, was used by Buras (1965) to find the optimum operating policy for the conjunctive use of a surface and a groundwater storage for irrigation. The state variables were the volume in storage in the surface reservoir, the volume in transit to replenish the groundwater storage and the volume in the groundwater storage. The stochastic variable was the inflow to the surface storage.

Decisions were required annually on the volume of water to release from the surface storage for irrigation and the volume for groundwater replenishment, and the volume to pump from the groundwater storage for irrigation to ensure that the net present value of returns was maximized.

The algorithm was solved on an annual basis. To enable this solution, coarse grids for the state and decision variables had to be used, because computer storage requirements increase in a geometrical manner with the number of state variables. Three state variables are often quoted as constituting a practical limit for direct dynamic programming solutions.

When operation over monthly intervals is considered, successive inflows will not normally be independent, but will display a persistence effect, measurable by the serial correlation coefficient.

Thus, if $p_{ij}(q_i / q_j)$ is the conditional probability of a flow q_i in the nth time period following a flow q_j in the (n+1)th time period,

$$f_n(s) = \max \left(R(d,s) + \frac{1}{1+r} \sum_{i=1}^{m} \left(\sum_{j=1}^{m} p_{ij}(q_i / q_j) \cdot f_{n-1}\left(s-d-L(s,d) + q_i\right) \right) \right), \qquad (1\text{-}10)$$

subject to

$$0 \leqslant d \leqslant \min\left(\left(s + q_i - L(s,d)\right), D_M\right).$$

$$0 \leqslant s \leqslant S_M$$

$$\sum_{i=1}^{m} p_{ij} = 1, \quad \text{for all } j, \quad 1 \leqslant j \leqslant m$$

Little (1955) used this type of stochastic algorithm to determine the optimal water storage policy for a hydro-electric power station. This solution worked on a two-weekly basis, assumed a known seasonal power demand pattern, and used as its objective the minimization of costs of alternative thermal generation.

With an approximate method of handling the conditional inflow probabilities, Schweig and Cole (1968) derived monthly operating rules for two interconnected water supply reservoirs using a stochastic algorithm.

1.7 OPTIMAL POLICIES BY DETERMINISTIC DYNAMIC PROGRAMMING

Although a number of notable solutions has been obtained using stochastic algorithms, the method is very demanding in terms of computer time and capacity, especially if conditional probabilities and a reasonable number of possible values for the state and decision variables are used.

An alternative, suggested by Hall and Howell (1963), is to return to the deterministic approach (equation (1-7)), find optimum patterns of releases for known sets of inflow data and combine these release patterns into a mean policy using a regression approach. For a system, this policy derived by use of the deterministic algorithm should, within the limits of sampling errors, be similar to that obtained from a stochastic solution.

Historic streamflow records are normally restricted in their length; few flow records of more than 50 years duration are available. Thus a mean policy derived from a deterministic solution using only a historic set of data may fail to take account of possible future sequences of inflows not represented in the historic record. Neither are any sequences of data left with which the real-time operation of the system using the mean policy may be tested against the absolute optimum policy obtained from a direct deterministic solution for a particular data set.

To overcome the first problem, Hall and Howell suggested that simulated sets of flow data, which resemble the historic data in terms of certain statistical parameters, could be used as input to derive the mean policy.

These simulated data sets can also be used to examine the closeness between real-time operation with this mean policy and the absolute optimum obtained from a direct deterministic solution.

The steps involved in this deterministic solution, which combines simulation and mathematical programming in a regression environment to solve a complex stochastic optimization problem, are illustrated as a block diagram in Fig. 1.3.

The number of repetitions of the simulation and optimization loop is not defined from within the problem, but by the level of confidence desired in the solution.

1.8 PREVIOUS INVESTIGATION OF THIS APPROACH

A preliminary study of this approach is given by Young (1966). The following are important points which arise from this investigation:

(1) Economic loss functions were used to value different levels of release. A target release was specified and economic losses were incurred if releases deviated from this target. The objective was to minimize these losses, the converse to maximizing economic benefits.

(2) Multiple regression was used to give best fit functions relating the optimum release d to the initial storage s and the inflow q during the ith period. Thus the mean policy was expressed in the best fit form,

$$d = f(s,q) \tag{1-11}$$

For economic losses as a quadratic function of d, linear policy functions were found to provide as good a fit as more complicated quadratic and cubic functions.

(3) The mean policy was compared with the classic standard release policy. This standard policy, illustrated in Fig. 1.4, states that if possible the target release should be made during the current period. By not allowing a small deficit in the current period to reduce to a possible much larger deficit in the future, this policy does not permit any hedging on releases. Any overflow which occurs when the storage becomes full is counted as part of the release; in Young's study releases in excess of the target incurred a penalty.

Figure 1.3

POLICY DERIVATION USING DETERMINISTIC OPTIMIZATION
(Roefs, 1968)

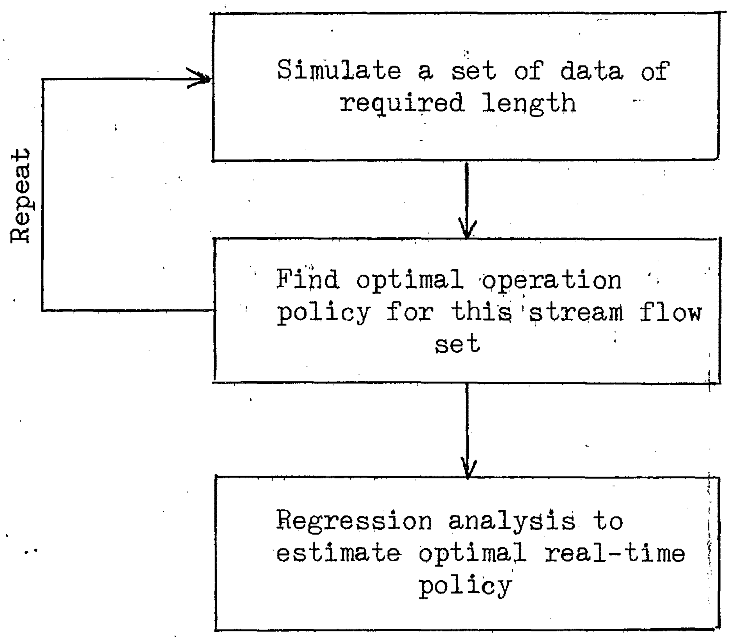
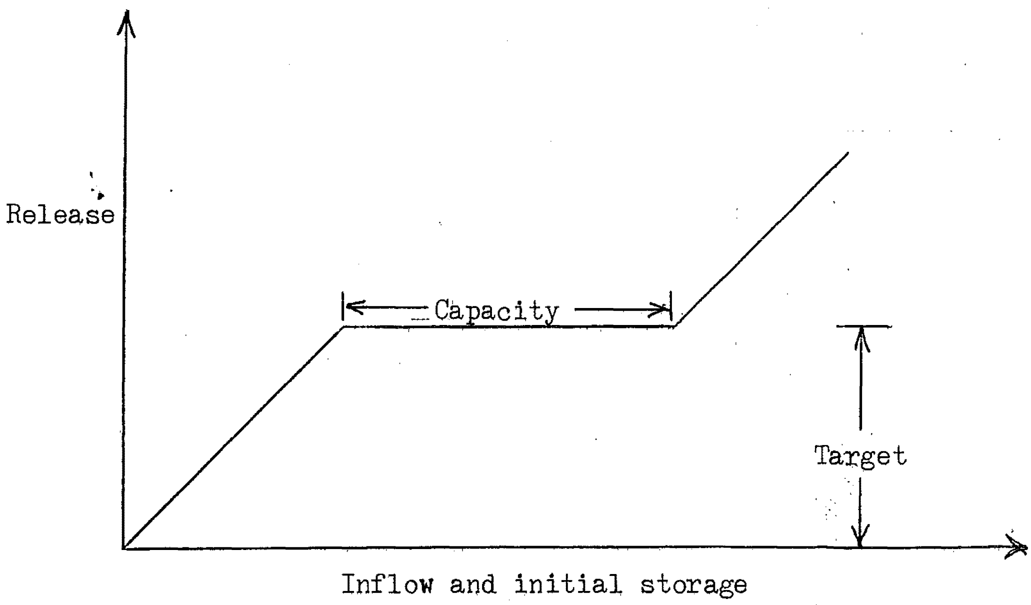


Figure 1.4

STANDARD POLICY



For some piece-wise linear loss functions, the standard policy was optimal.

(4) The deterministic algorithm optimizes over a finite time horizon and water remaining in storage at the end of this period is of no value. As this end point is approached the optimizing problem progressively loses its steady state characteristics and gradually becomes a terminal state problem. Results were given defining the time over which this terminal effect is significant.

(5) The deterministic algorithm was shown to be more efficient in terms of computational requirements than the stochastic dynamic programming algorithm. The name "Monte Carlo Dynamic Programming" (MCDP) was coined for the deterministic algorithm.

(6) An ability to use outside information to forecast inflows one period ahead was shown to be of economic merit in certain cases.

The following factors limit the applicability of the results of this study to practical storage operation problems:

(1) The best fit policy function (1-11) used the current inflow q as an independent variable. The use of this best fit policy in the real time operation of a storage requires the inflow q during the ith period to be known before the release decision d can be made. In many cases, especially when the time intervals are less than one year, the decision d must be made before the volume of current inflow is known.

(2) All spills from the storage were presumed to be part of the release from which the economic loss was evaluated. If the objective in operation is to maintain a steady flow pattern downstream of the storage this is satisfactory, but in certain water uses, notably water supply, irrigation and hydro-power generation, the volume of spillage does not directly affect the magnitude of benefits.

(3) The added complications resulting from within year seasonal variations both in inflows and targets were not touched on.

(4) No definition of the required length of deterministic solution was given.

(5) No interest rate was included. Implicitly, the value of a lump sum of money received now or in the future did not change.

(6) No allowance was made for seepage and evaporation losses from the storage.

(7) Inflows to the storage were assumed to be normally distributed.

(8) It is not shown that the best fit policy functions gave good estimates of the true optima.

The versatility of the dynamic programming approach is such that these features can be readily handled within the proposed solution algorithm. That this is so will be demonstrated in the following chapters.

CHAPTER TWO

STREAMFLOW ANALYSIS AND SYNTHESIS I, THE SINGLE STREAM MODEL

2.1 INTRODUCTION

Streamflow is an example of a hydrologic process. Recorded data from most hydrologic processes give arrays of variables which tend to be non-random and non-stationary and which do not follow normal frequency distributions (Chow, 1964).

Thus those standard statistical techniques, which require a priori assumptions of randomness, stationarity and normality, must be used with care in the processing of hydrologic data. In addition, many hydrologic processes must be observed historically and cannot be measured in a controlled experimental environment. Thus observations available on a particular process may be few, limiting the extent of an analysis and the validity of conclusions which can be drawn from it.

Streamflow observations are non-random because of persistence effects in runoff from successive time intervals, and non-stationary because the statistics of flows in successive time intervals are not the same, but tend to follow seasonal patterns. Typically, sets of recorded flows tend to follow non-normal probability distributions. If one is concerned with monthly or annual flow volumes, the amount of data available is limited by the time during which recordings have been taken at the flow gauging station.

The term "stochastic process" refers to a "time dependent probabilistic process". Streamflow may be described as a non-stationary stochastic process.

The analysis of data on the streamflow process, and the development of a statistical model to represent it is an exercise in time series analysis.

All the theory of time series assumes stationary processes. The conversion of a non-stationary process into a stationary process requires the use of certain transformation functions. Often the conversion may be an iterative process, requiring the use of time series analysis to detect non-stationary effects and the use of transformation functions to eliminate them.

2.2 TIME SERIES

"A time series is a sequence of values arranged in order of their occurrence which can be characterized by statistical properties." (Chow, 1964).

Time series are involved in many fields of study. The major concern of this study is with streamflow, an example in hydrology. Although most hydrologic time series represent continuous processes, observations are made at intervals of time and the resulting series are discrete. With streamflow, this time interval is often one day.

In general, a time series comprises two parts, a stochastic component and a deterministic component.

2.2.1 Deterministic Component

In general, the deterministic component is a result of trend and cyclic effects, and these both tend to make a time series non-stationary.

Trend may be considered as a smooth motion in a series over a long period of time. Trend analysis is subject to uncertainty, since it is always possible for an apparent trend to be part of a slow oscillation covering a period of time much longer than the record. Trend may be present in streamflow records if development over a number of years has resulted in increasing consumptive withdrawals of water from the stream or if the catchment conditions have changed progressively. Julian (1961) showed that trend was present in the record for the Colorado River at Lees Ferry. Trend will be assumed absent from all the data used in this study.

Cyclical effects must be distinguished from oscillatory effects. In a cyclical component the maxima and minima occur at equal intervals of time and with constant amplitude, but are distorted by the stochastic component. In an oscillatory effect the intervals between maxima and minima and their amplitudes are randomly distributed. A cyclical time series is oscillatory, but an oscillatory effect is not necessarily cyclical.

2.2.2 Stochastic Component

The stochastic component remains as a time series after cycles and trends have been removed. In general the stochastic component contains both a residual time-dependent deterministic component and a pure random component.

2.3 TIME SERIES ANALYSIS

Two different points of view may be taken in analysing time series; from a basic viewpoint one would attempt to identify trend and cyclic effects and to relate them to other phenomena, while from an applied point of view one would seek to define a model of the process in terms of certain statistical parameters, with the intent of using this model to represent the process in a wider study. This applied viewpoint is adopted here. Note that this viewpoint is very dependent on the results obtained in previous basic studies.

Methods for detecting the presence and nature of deterministic components in time series include moving averages, variance spectrum analysis and serial correlation analysis. Variance spectrum analysis and serial correlation analysis are to some extent complementary in detecting significant cyclical patterns. The latter technique is used in this study.

2.3.1 Serial Correlation Analysis

The serial correlation coefficient is analogous to the product moment correlation coefficient for two sets of data. If the variable x_t, $(1 \leqslant t \leqslant N)$, represents a time series of length N, the kth order serial correlation coefficient r(k) may be calculated from -

$$r(k) = \frac{\sum_{t=1}^{N} x_t \cdot x_{t+k} - (\sum_{t=1}^{N} x_t)^2/N}{\sum_{t=1}^{N} x_t^2 - (\sum_{t=1}^{N} x_t)^2/N} \qquad (2\text{-}1)$$

Note that $r(0) = 1$,

and $-1 \leqslant r(k) \leqslant 1$ for all integer k, $k > 0$.

This is not an exact expression; rather it assumes a circular series in which x_1 follows x_N, but the resulting error in r(k) is negligible for $k \ll N$.

The behaviour of the serial correlation as k increases successively from unity reflects the nature of the series. Values of this function plotted from successive k values define a correlogram, a useful device in time series analysis.

A correlogram enables cyclical components to be identified, as well as assisting in the division of the stochastic component into non-random time dependent and pure random components. A line can be drawn through successive points on a correlogram to illustrate better the behaviour of r(k) with increasing k. This line serves no other purpose, since only the plotted r(k), corresponding to integral values of k, are of interest.

An estimate of the serial correlation coefficient is made from a sample of N observations of a theoretically infinite population. For this reason, r(k) is subject to sampling errors from the true population serial correlation coefficient $\rho(k)$. To distinguish the sample estimate from the population value, $\rho(k)$ is often referred to as the kth order auto-correlation coefficient (Matalas, 1967a).

As sample estimates, the values of r(k) may be non-zero even when the population value $\rho(k)$ is zero. A widely used approximate test for the significance of non-zero values is given by Matalas (1967a). Confidence limits on r(1) are:

$$\text{C.L.}\ (r(1)) = (-1 \pm z_\alpha\ (N-2)^{\frac{1}{2}}) / (N-1) \qquad (2\text{-}2)$$

where z_α is a normal deviate corresponding to the desired level of confidence α.

If $r(1)$ is outside the confidence limits, the hypothesis that the auto-correlation coefficient $\rho(1) = 0$ is rejected. This test can be extended to test the significance of $r(k)$ for $k > 1$ and $k \ll N$.

As k increases, sampling errors of $\rho(k)$ increase; normally the maximum lag for calculating $r(k)$ is in the range N/10 to N/4.

On a correlogram cyclical components are evidenced by regular cyclic movements. In streamflow data, one might expect cyclical effects of an annual period. There is too the possibility of periods greater than one year. A great deal of effort has been directed towards determining long term cycles in hydrologic time series, but few encouraging results have been obtained. Records of 40 to 60 years length are too short to enable a clear distinction to be made between long term "hidden periodicities" and oscillatory patterns.

Before further analysis can proceed, significant cycles detected on the correlogram must be removed from the series. A convenient way to obtain a stationary time series from a monthly flow series which contains an annual cycle is to use the transform,

$$y_t = \frac{q_t - \bar{Q}_j}{S_j} \qquad (2\text{-}3)$$

for all t and j, $1 \leqslant t \leqslant 12 * N$, $1 \leqslant j \leqslant 12$,

where; y_t is a stationary stochastic series with zero mean and unit variance.

N is the length of the record in years.

t is the number of months from the start of the record.

j is the position of the t th month in the 12 month annual cycle.

q_t is the flow volume for the t th month.

$\bar{Q}_j$ is the average flow for the jth month of the 12 month annual cycle.

S_j is the standard deviation of all flows in the jth month.

This transform results in a new stationary stochastic y_t series, with zero mean and unit variance. An alternative harmonic analysis method, useful for obtaining stationary series for daily flow data, is given by Quimpo (1967). The y_t series may be referred to as a residual series.

2.3.2 Autoregression Processes

Once a stationary series has been obtained, several types of mathematical generating functions may be used to represent the series. Frequently used in hydrologic analyses are linear autoregression processes. A general linear autoregression process is -

$$y_t = \sum_{k=1}^{p} \alpha_k \, y_{t-k} + \epsilon_t, \tag{2-4}$$

where; y_t represents a stationary stochastic series with zero mean and unit variance.

p is the order of the process

α_k are coefficients to be determined

ϵ_t is a pure random series.

2.3.3 First Order Markov Series

If $p = 1$, (2-4) becomes

$$y_t = \alpha_1 \, y_{t-1} + \epsilon_t. \tag{2-5}$$

This is the simplest case of an autoregression series and is known as a first order Markov, or Markov lag-one series. In this series,

$$\alpha_1 = \rho(1), \tag{2-6}$$

where $\rho(1)$ is the first order autocorrelation coefficient. Thus,

$$y_t = \rho(1) \; y_{t-1} + \epsilon_t. \tag{2-7}$$

The variance of the random series, ϵ_t, is given by

$$\begin{aligned} \text{var}\,(\epsilon_t) &= (1 - \rho(1)^2).\ \text{var}\,(y_t), \\ &= 1 - \rho(1)^2, \end{aligned} \tag{2-8}$$

since $\text{var}\,(y_t) = 1.$

Persistence, the influence of past values of y_t on the present value, is represented by the y_{t-k} term in (2-4). In a Markov lag-one process, the history of the process is summed up in y_{t-1}, as in (2-5). A hydrologic reason for persistence in time series of streamflows is the storage of water in the catchment, either in lakes, or as groundwater or snow.

For a lag-one process, the lag-k autocorrelation coefficient $\rho(k) = \rho(1)^k$.

Thus, the theoretical correlogram for a lag-one process assymtotically approaches zero as k increases, since $\rho(1) < 1$.

2.3.4 Second Order Markov Series

With p = 2, equation (2-4) becomes

$$y_t = \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \epsilon_t \qquad (2\text{-}9)$$

This is a second order linear autoregression or Yule series. α_1 and α_2 can be obtained from the relationships,

$$\left.\begin{aligned} \alpha_1 &= \frac{\rho(1)\ (1-\rho(2))}{1-\rho(1)^2} \\ \alpha_2 &= \frac{\rho(2)-\rho(1)^2}{1-\rho(1)^2} \end{aligned}\right\} \qquad (2\text{-}10)$$

The variance of ϵ_t is given by,

$$\frac{\mathrm{var}\ (y_t)}{\mathrm{var}\ (\epsilon_t)} = \frac{1-\alpha_2}{(1+\alpha_2)\ ((1-\alpha_2)^2-\alpha_1^{\ 2})}, \qquad (2\text{-}11)$$

(Kendall and Stuart, 1966).

The lag k autocorrelation coefficient is given by,

$$\rho(k) = \frac{(-\alpha_2)^{k/2} \sin(k\Theta + \varphi)}{\sin \varphi} \qquad (2\text{-}12)$$

where

$$\Theta = \cos^{-1} \frac{\alpha_1}{2\,(-\alpha_2)^{\frac{1}{2}}},$$

and $\tan \varphi = \dfrac{1+\alpha_2}{1-\alpha_2} \tan \Theta$.

This correlogram is a harmonic function with frequency Θ damped by a factor $(-\alpha_2)^{k/2}$.

2.3.5 Test of Fit of Autoregression Process

In order to test whether an observed series y_t can be represented by an autoregression process, a new series Z_t can be produced by

$$Z_t = y_t - \sum_{k=1}^{p} a_k \, y_{t-k}, \tag{2-13}$$

where a_k are estimates of the autoregression coefficients α_k. This Z_t series can be checked for serial independence by calculating its correlogram. Where serial independence is shown for this series, the y_t series is accepted as coming from an autoregression process of order p.

To verify the fit of the scheme, the calculated variance of Z_t can be compared with its theoretical value. For first order and second order schemes, the theoretical variance is given by (2-8) and (2-11) respectively.

For a first order scheme, p = 1, and

$$Z_t = y_t - a_1 \, y_{t-1}. \tag{2-14}$$

a_1 is an estimate of α_1, and r(1) an estimate of $\rho(1)$; therefore, from (2-8), $a_1 = r(1)$.

For a second order scheme, p = 2, and

$$Z_t = y_t - a_1 \, y_{t-1} - a_2 \, y_{t-2}. \tag{2-15}$$

Again, using estimates of population variables in (2-10),

$$\left.\begin{aligned} a_1 &= \frac{r(1) \, . \, (1 - r(2))}{1 - r(1)^2} \\ \text{and} \quad a_2 &= \frac{r(2) - r(1)^2}{1 - r(1)^2} \end{aligned}\right\} \tag{2-16}$$

2.3.6 The Log-Normal Distribution

To completely specify the time series model of the streamflow, a probability distribution must be assumed for the random series. A number of different distributions have been used to model streamflow data. In general, monthly and annual series tend to be positively skewed, and skewed probability distributions are required.

The log-normal probability distribution is widely used in hydrologic analyses. As its name suggests, this distribution takes the logarithms of a series as being normally distributed. Thus, if this distribution is appropriate for a q_t series, the $\log_e q_t$ series would be expected to be normally distributed.

Given a q_t series, 5 different methods are available for estimating the mean μ and the standard deviation σ of the $\log_e q_t$ series, (Aitchison and Brown, 1957). Their criteria for choosing a method are -

(1) that it be unbiased (consistent),

(2) that it minimize the variance of the estimates, and

(3) that the calculations involved be reasonably compact.

Two of the methods are described below.

The maximum likelihood method is theoretically the best possible and simply involves estimating μ and σ from the $\log_e q_t$ series with the equations-

$$m = \frac{1}{n} \sum_{t=1}^{n} \log_e q_t$$

and $$s^2 = \frac{1}{n-1} \sum_{t=1}^{n} (\log_e q_t - m)^2,$$

where m and s are estimates of μ and σ.

This method is unwieldy to use manually if n is large.

The method of moments, or the moment transformation method, gives m and s directly from the mean $\bar{Q}$ and the standard deviation S of the q_t series by the relationships -

$$m = \log_e \bar{Q} - \tfrac{1}{2} \log_e \left(1 + \left(\frac{S}{\bar{Q}}\right)^2\right) \tag{2-17}$$

and $$s^2 = \log_e \left(1 + \left(\frac{S}{\bar{Q}}\right)^2\right) \tag{2-18}$$

If these are solved for m and s, the values obtained will in general be different to the maximum likelihood estimates.

Although this method of moments is widely used in hydrology for estimating parameters for log-normal distributions, Aitchison and Brown do not advocate it, noting that "the method of moments has little to recommend it either theoretically or computationally (especially when σ is large)." Their preference is for maximum likelihood estimates.

2.3.7 Distribution of the Z_t Series

In this study of monthly and annual flows, interest is in approximating with theoretical distributions the probability distribution of the random Z_t series. The parameters of the Z_t series can be specified as functions of the standardized y_t series.

If $\bar{y}_t = 0$, $\bar{Z}_t = 0$ also. If a first order Markov model has been assumed, var (Z_t) can be obtained from (2-8), thus,

$$\text{var}\,(Z_t) = (1 - \rho(1)^2) \cdot \text{var}\,(y_t).$$

Again for a first order model, γ_z, the skewness of Z_t, is related to the skewness of y_t by the expression, (Fiering, 1967)

$$\gamma_z = \gamma_y \cdot (1 - \rho(1)^3) \,/\, (1 - \rho(1)^2)^{3/2} \tag{2-19}$$

If y_t is normally distributed, $\gamma_y = 0$; consequently $\gamma_z = 0$ and Z_t is also normally distributed.

If y_t is skew distributed, $\gamma_y \neq 0$, and $\gamma_z \neq 0$. Because of the effect of serial correlation, $\gamma_z \neq \gamma_y$, unless $\rho(1) = 0$.

Summing up, if the parameters of y_t are known and a first order model is assumed, the expected parameters of Z_t can be derived. In particular, if y_t follows a normal distribution, Z_t is also expected to follow a normal distribution.

2.3.8 Monthly Series

Difficulties can arise in that skewness coefficients for different months and the covariance function for successive pairs of monthly data may not have the same expected value for all months. This would imply that the series is non-stationary in terms of skewness and covariance respectively.

2.3.9 Third Order Stationarity

If the random series is normally distributed, it can be described as strictly stationary, since stationarity of all moments of order greater than two is implied by a normal distribution. For a monthly series second order stationarity is achieved by the transformation (2-3).

Use of the log-normal frequency distribution implicitly assumes third and higher order stationarity for the transformed series. With a monthly series this is not strictly valid, but large sampling errors with higher order moments tend to mask possible patterns of month-to-month variation in the skewness coefficients and higher order parameters.

2.3.10 Covariance Stationarity

Use of the autoregression process presumed an autocorrelation structure which was the same between each pair of months. This results in a single lag-one serial correlation coefficient r (1).

An alternative definition implies 12 lag-one serial correlation values r_j, j = 1, 2 .... 12, wher r_j is the correlation between flows of the jth and (j-1)th months. (If j = 1, read (j - 1) = 12, since r_1 is correlation between flows in the 1st month of each year and flows in the last month of the previous year). The mean of the 12 r_j = r (1), thus in a sense r (1) may be described as an average serial correlation coefficient.

If the 12 r_j are not significantly different to each other, or do not show seasonal tendencies, the process may be described as covariance stationary. There are two reasons for believing that monthly flows may not be covariance stationary;

(1) in rivers subject to seasonal runoff patterns, dry season flows, consisting largely of base flow, might be expected to show higher serial correlations than wet season flows, which are subject to perturbations from rainfall (Moreau, 1968),

(2) changes in the time and duration of the spring thaw might account for weak correlations observed between flows in spring and early summer months, (Thomas and Fiering, 1962).

If the average value is quite large and month to month patterns are judged to be significant, then use of a covariance stationary model to represent the series would result in a distorted representation.

Nevertheless, it will be demonstrated in this chapter that sampling errors in the estimation of the covariance function are large and as with the skewness coefficient, tend to mask possible patterns of variation. For this reason, the use of one average serial correlation value instead of 12 monthly values may well be a satisfactory approximation.

2.3.11 Review of Applications

Ideally, a correlogram will by the nature of its shape indicate the type of generating process from which the series derive. In practice, there are some difficulties. Sampling fluctuations tend to obscure the behaviour of the correlogram as k increases. Undetected and ill defined cyclical components may add to the fluctuations. The series may actually be a linear combination of several different schemes.

Despite these difficulties, lag-one autoregression processes have been extensively used in hydrology to model annual and monthly flow series (Julian, 1961; Yevdjevich, 1964; Roesner and Yevdjevich, 1966).

Yevdjevich (1964) showed that the mean first order serial correlation coefficient for a worldwide set of 140 long term annual runoff records was 0.17 with standard deviation 0.18. Values of r(1) ranged from -0.348 to 0.705. Large values of r(1) such as the 0.70 obtained for 97 years record for the St Lawrence River were ascribed to storage effects of the catchment, in this case the Great Lakes. Negative values of r(1) significantly different to zero imply a tendency for high values to follow low values, and low values to follow high values. Such a situation is difficult or impossible to justify hydrologically.

Some persistence in runoff may be due to persistence in seasonal precipitation volumes, but most rainfall records do not show significant persistence (Roesner and Yevdjevich, 1966). In general, a stationary time series for seasonal rainfall volumes is a pure random series. Roesner and Yevdjevich (1966) found that linear lag-one Markov models fitted the majority of a large sample of monthly runoff records.

Quimpo (1967) showed that a second order linear Markov model was required to represent a time series of daily streamflow volumes.

2.4 SYNTHETIC HYDROLOGY

Given a time series model of the flow process, and an appropriate frequency distribution for the random component, the reverse of the analytic procedure can be used to prepare further "synthetic" time series.

By sampling from the appropriate distribution a pure random series can be constructed. With an arbitrary y_1 value, a y_t series can be simulated using an autoregression process of the type (2-4). Finally the y_t can be transformed by the inverse of (2-3) to form a "synthetic" flow sequence.

In terms of the statistical parameters specified in the process, such a synthetic sequence should be indistinguishable from the original historic data series. The technique outlined here is known under a variety of headings but the term synthetic hydrology is in wide usage.

As many synthetic sequences as required and of any length may thus be generated. Each of these sequences represents, in terms of the statistical parameters preserved, possible sequences of streamflow. Each is as likely to occur as any other. On this reasoning, a well balanced design of a water resource system will reflect the experience of many such sequences.

2.4.1 Bias

The parameters $\bar{Q}_j$, S_j, $r(1)$ and the skewness C_s of the historic sequence are sample estimates of the population values. From a statistical point of view each is an unbiased and consistent estimate. These estimates are unlikely to equal their respective population values; rather they will vary about the population values with a magnitude dependent on their respective errors of estimate.

Synthetic sequences generated will not preserve the population values of the parameters, but rather the historical sample estimates; thus the expected value of the synthetic parameters will not be the population values, but the sample estimates. With respect to the synthetic sequences, the historic estimates are biased.

Bias cannot be eliminated, because the true population values are unknown. It can be minimized by using correct methods for estimating sample parameters, and possibly by pooling regional information to estimate parameters at a particular station.

In addition to the four classes of parameters described above, others could be used to characterize the historic sequence. These additional parameters would involve higher order moments and time lags, as well as increased errors of estimate. Bias increases with the error of estimate and the use of these extra parameters in a generating algorithm is probably of doubtful value. (Matalas, 1967a).

2.4.2 Literature Review

Although the name synthetic hydrology is new, the idea is not. Early efforts at streamflow synthesis were directed at annual flows.

To produce an extended record for a single stream, Hazen (1914) combined successively the flow records for 14 adjacent stations, scaling each record to give it the same mean flow as the station under consideration. The combined records from 14 streams gave a 300 year record.

Sudler (1927) took 50 cards and printed a representative annual flow on each. The deck was shuffled and dealt 20 times over. In this way a 1000 year synthetic record was obtained.

Barnes (1954) randomly sampled from a normal distribution of appropriate mean and variance to generate a synthetic flow sequence. The random sampling was effected by selecting from the normal distribution 100 flow values, all of equal probability of occurrence, and assigning a code number to each. As code numbers appeared on a random number table, appropriate flows were recorded. In this way a 1000 year sequence was obtained.

These techniques were for generating annual data for stations where records did not contain significant persistence effects. Such is the case in many smaller rivers.

When attention is shifted to monthly data, persistence is nearly always present and cannot be disregarded. A generating model to cope with this can be derived in the following way.

Combining equations (2-3) and (2-14) to eliminate y_t and y_{t-1},

$$Z_t = \frac{q_t - \bar{Q}_j}{S_j} - r(1) \frac{q_{t-1} - \bar{Q}_{j-1}}{S_{j-1}} \qquad (2\text{-}20)$$

Defining t_t as a pure random series with zero mean and unit variance, and the same distribution as Z_t,

$$Z_t = (1 - r(1)^2)\, t_t,$$

$$\text{since } \operatorname{var}(Z_t) = 1 - r(1)^2.$$

Rearranging (2-20)

$$q_t - \bar{Q}_j = \frac{S_j}{S_{j-1}} .r(1).(q_{t-1} - \bar{Q}_{j-1}) + t_t.S_j.(1-r(1)^2)^{\frac{1}{2}} \qquad (2\text{-}21)$$

If allowance is made for seasonal variations in the serial correlation,

$$q_t - \bar{Q}_j = \frac{S_j}{S_{j-1}} .r_j(q_{t-1} - \bar{Q}_{j-1}) + t_t.S_j.(1-r_j^2)^{\frac{1}{2}}, \qquad (2\text{-}22)$$

where r_j is the serial correlation between flows in the jth and (j-1)th months of the 12 months annual cycle. (If j = 1, read j-1 = 12.)

Equation (2-22) was described by Thomas and Fiering (1962), and has become known as the Thomas-Fiering equation. It appears to represent the first application of linear autoregression type models to monthly flow data.

To generate a sequence of synthetic flows, a value q_o is assumed to start the process. A random number t_1 is selected and a value for q_1 in the first month is calculated from (2-21) or (2-22). A second random number is then used with q_1 to generate a flow for the second month, and so on. Once started, the generating procedure can be used to produce a synthetic record as long as required.

If in (2-21), the subscript j takes the value 1 only, the equation becomes a one season annual flow model of the type used by Brittan (1961).

A number of additions to the basic model have appeared. Hufschmidt and Fiering (1966) showed that (2-22) can be applied when monthly flows are neither normally nor log-normally distributed. Where monthly flows follow a gamma distribution, the generating equation will give correctly skewed flows if the standardized random deviates t_t have a skewness dependent upon, but not equal to, the skewness of observed monthly values. This relationship is given by (2-19). A transform suitable for changing normally distributed standard deviates to standard deviates with the appropriate skewness is available (Fiering, 1967).

Harms and Campbell (1967) introduced an addition to the model which preserved the observed serial correlation between annual flows, as well as the relationship between monthly flows.

Matalas (1967b) compared the use of log-normal and gamma distributions. Though the error involved is small, maintaining the serial correlation between the logarithms of flows does not guarantee the maintenance of serial correlations in non-transformed values. This paper gave an exact solution which involved modifying the serial correlation between the natural values.

Young and Pisano (1968) illustrated the use of what they called a <u>minimum skewness transform</u>. After obtaining second order stationarity for monthly data, they calculated the sum of skewness coefficients for 12 monthly sets of the stationary y_t series. They repeated the procedure;

(a) for logarithms of monthly flows and

(b) for square roots of monthly flows.

The logarithmic transform most often gave minimum total skewness.

A multilag model was described by Fiering (1967). Expressions were developed relating serial correlation coefficients in runoff to auto-correlation coefficients for a Markov rainfall process. Persistence in runoff was postulated to be due to persistence in precipitation and to carryover effects of groundwater storage.

Recent studies have questioned the use of linear autoregression type models, because they do not always model well the observed extremes of flood and periods of drought. Fiering (1967) suggested the use of multi-lag models, while Mandlebrot and Wallis (1968) introduced a new class of "self similar" models which use fractional Gaussian noise time series. These MW (Mandlebrot and Wallis) models, aimed at specifically modelling low frequency characteristics of time series, should reproduce critical periods of runoff better than linear autoregression models.

It may be that only a very small improvement is to be gained by using the computationally demanding MW models in preference to autoregression models to simulate data for design purposes. On the other hand, a design based on synthetic data which does not contain critical periods as severe as those in the historic data might well give a less conservative design than that obtained using the historic data. This can also be interpreted as an inadequacy of the generating technique.

To some extent, linear autoregression models may be improved by using a log-normal distribution for the flow data. A log-normal distribution emphasizes low flow volumes compared to high volumes and at least in some casea a log-normal autoregression model should generate critical periods closer to the historic than those generated by normal autoregression models.

For a sample of 30 different monthly flow records, Askew et al (1970) observed that a log-normal lag-one autoregression (Markov) model or a lag-one model with logarithms following a Pearson distribution reproduced critical periods of runoff better than other lag-zero and lag-two normal and log-normal autoregression models. On this basis it was concluded that lag-one log or log-Pearson autoregression models were superior to other autoregression models at generating critical periods similar to those observed historically.

However, this was hardly a complete analysis. It is likely that co-variance non-stationary models could generate critical periods much closer to the historic critical periods than could the covariance stationary models used in this study, but no mention was made of covariance non-stationary models.

It was shown that at least in some cases, MW models gave better critical periods than covariance stationary autoregression models.

TABLE 2.1

ANNUAL FLOW STATISTICS FOR SEVEN NEW ZEALAND STATIONS

Station	Ref. No.	Length Record (yrs)	Annual Flow Statistics: Mean CSD x 1000	Std dev. CSD x 1000	Skewness coef.	Lag-one serial correl.
L. Tekapo inflows	1	40	1078	176	0.28	0.06
L. Pukaki inflows	2	32	1661	270	0.45	-0.07
L. Ohau outflows	3	39	981	195	0.91 *	0.04
Ahuriri R. at Benmore	4	15	394	74	1.41 *	0.21
L. Rotoiti outflows	5	59	280	46	0.55	0.42 *
L. Manapouri outflows	6	35	5044	689	0.75	-0.10
L. Te Anau outflows	7	35	3576	530	0.82 *	-0.04

* Significant at 95% level

2.5 DATA ANALYSIS

The monthly flow series for seven flow recording stations in New Zealand have been analysed. All but one of these are in fact inflows or outflows from lakes which have been considered for hydro-electric power generation.

The stations are listed in Table 2.1 where they are assigned reference numbers. Subsequent reference to these stations will use these numbers.

Table 2.1 includes the annual flow statistics for the flow series belonging to these stations. For station 5 only, the serial correlation is significant, whilst skewness is significant for 3, 4 and 7.

Average monthly hydrographs are shown in Figure 2.1. An indication of the variability of the data is given by plotting $\bar{Q}_j \pm S_j$. If flows in a particular month were normally distributed, about 2/3 of the flow observations for this month would be expected to be within this range.

In the references to months in this thesis, month 1 represents January, month 2 represents February etc. Note that the Southern Hemisphere winter

FIG. 2.1 AVERAGE MONTHLY HYDROGRAPHS ±ONE STANDARD DEVIATION (in CSD x1000)

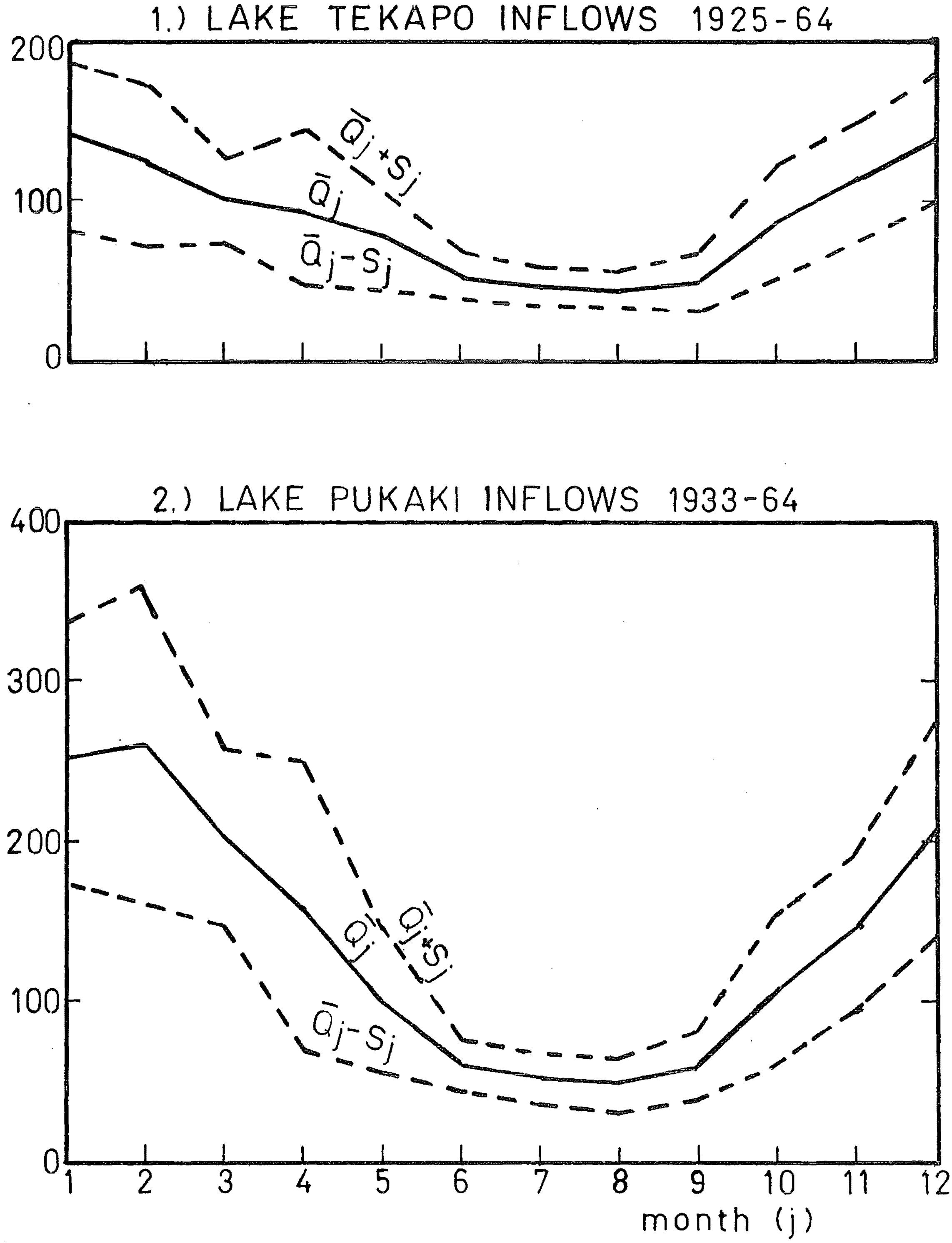


FIG. 2.1(contd)

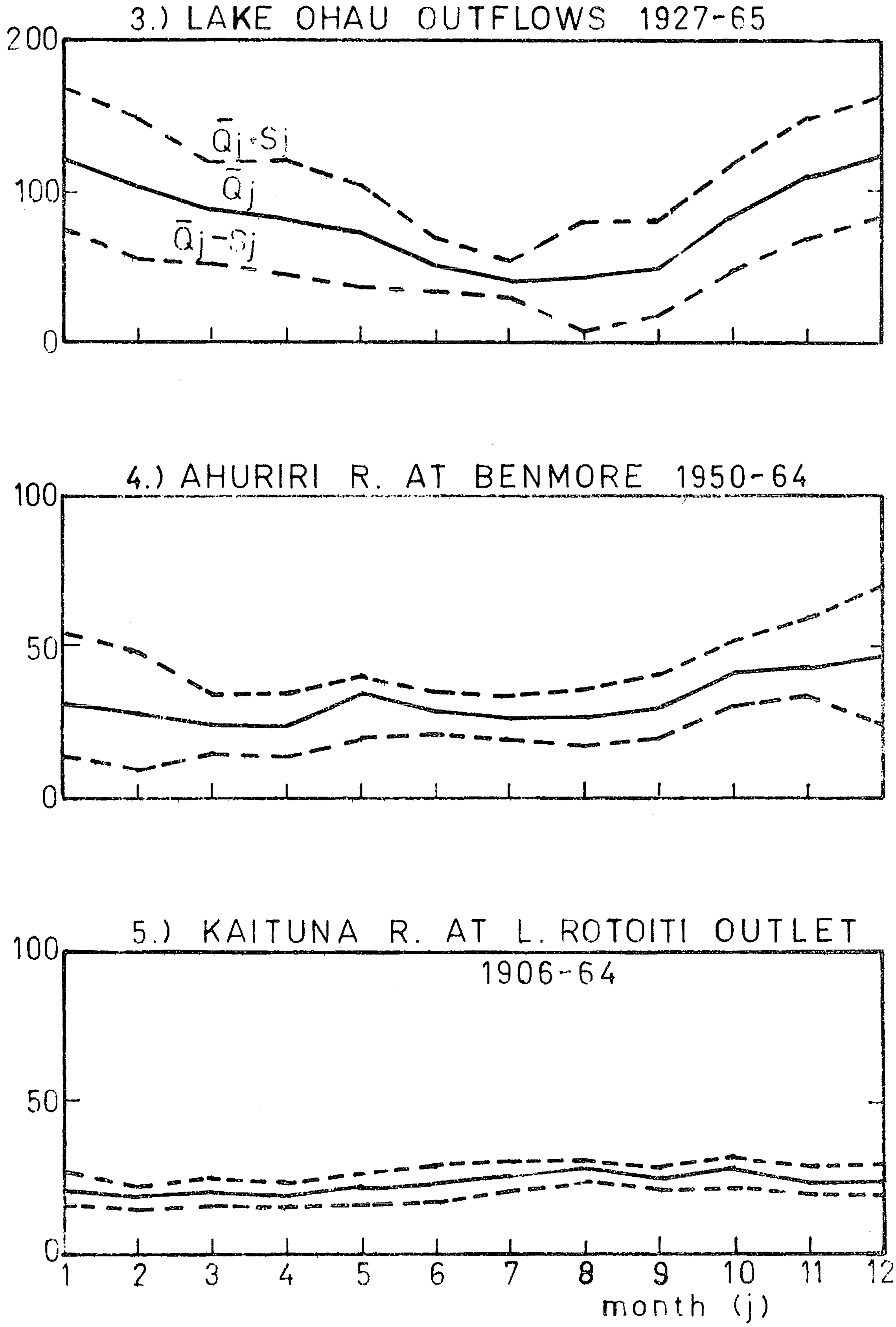
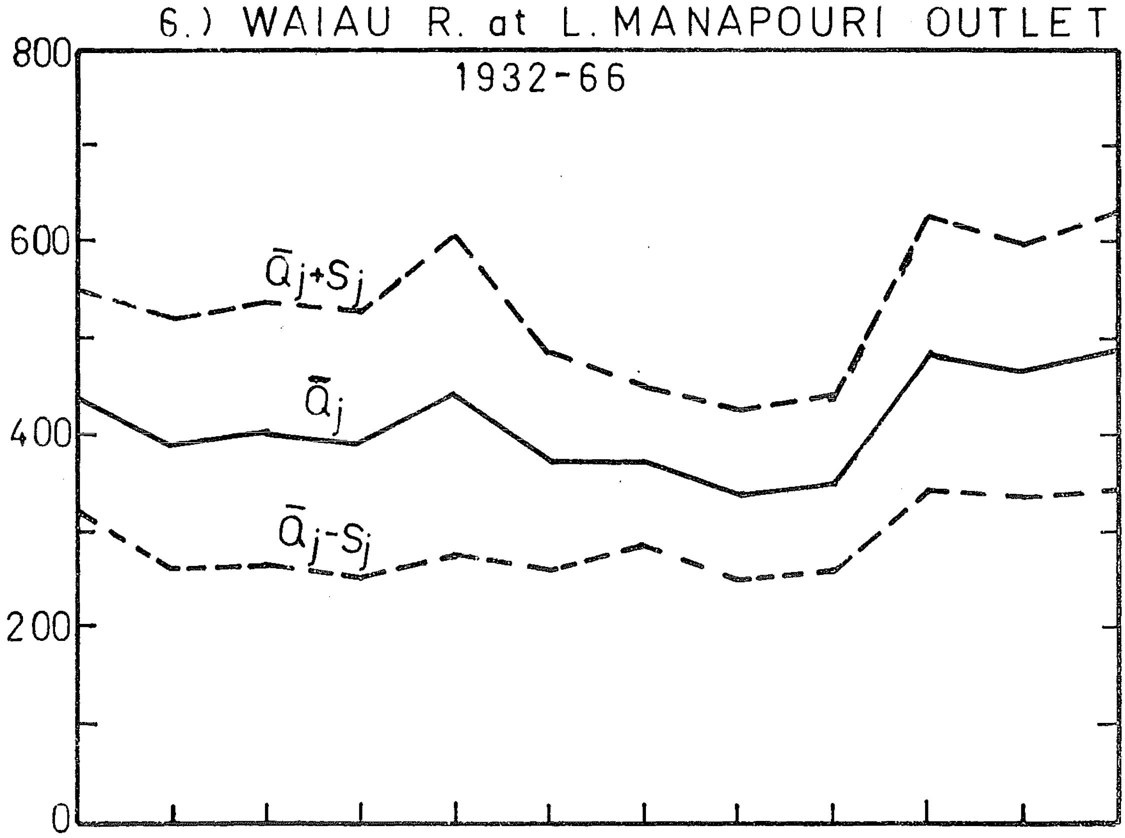
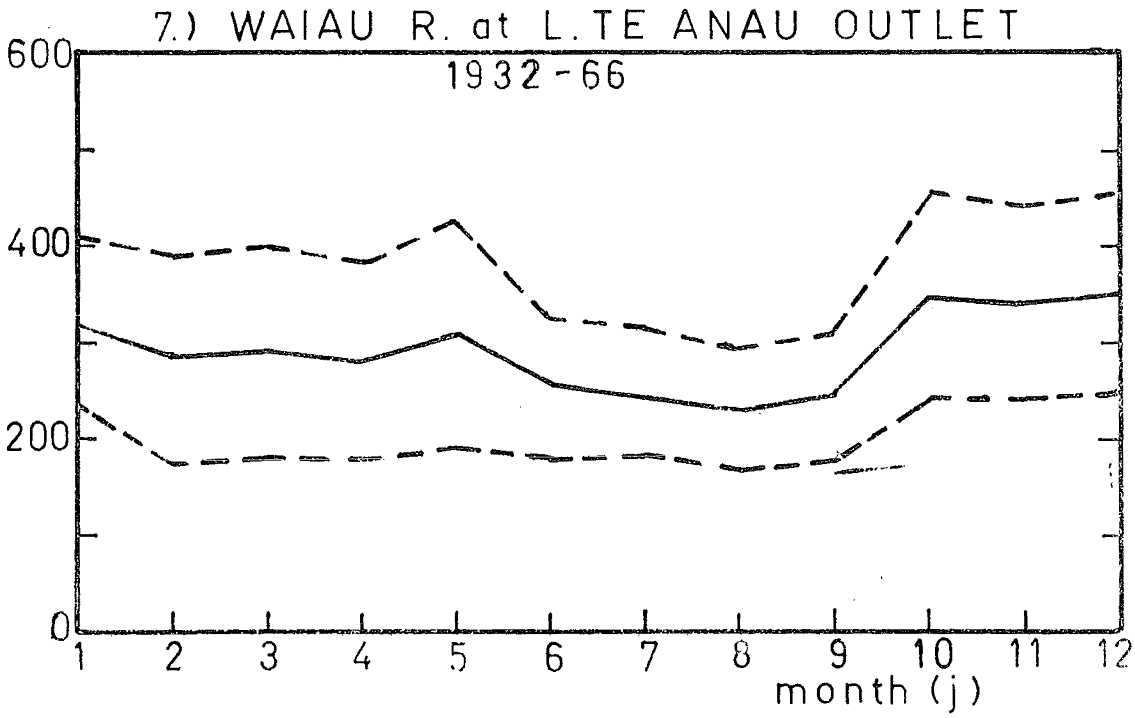

3.) LAKE OHAU OUTFLOWS 1927-65
200
100
0
$\bar{Q}_j + S_j$
$\bar{Q}_j$
$\bar{Q}_j - S_j$
4.) AHURIRI R. AT BENMORE 1950-64
100
50
0
5.) KAITUNA R. AT L. ROTOITI OUTLET
1906-64
100
50
0
1 2 3 4 5 6 7 8 9 10 11 12
month (j)

FIG. 2.1 (contd)


6.) WAIAU R. at L. MANAPOURI OUTLET
1932-66
800
600
400
200
0
$\bar{Q}_j+S_j$
$\bar{Q}_j$
$\bar{Q}_j-S_j$


7.) WAIAU R. at L. TE ANAU OUTLET
1932-66
600
400
200
0
1
2
3
4
5
6
7
8
9
10
11
12
month (j)

occurs over months 6, 7 and 8 while summer is over months 12, 1 and 2.

The first three records illustrated in Figure 2.1 are all from stations in the Waitaki River Basin in the South Island. Summer flows, which are high and variable in comparison to winter flows, originate in part from snow and ice melt.

Plots of month to month serial correlations are given in Figure 2.2. For stations 1, 2, 3 and 4, these r_j values show wide variability. Lower than average values are frequent for months 1, 4, 6, 7 and 10, while higher than average values are frequent for months 3 and 9. About all that can be inferred from this is that there tends to be a band of lower than average values for months 4 to 7, in the autumn and early winter.

For stations 1, 2 and 3, the average values, equivalent to the respective $r(1)$ values, are rather low. The average for station 4 is greater, but individual values show wide variability. These values are estimated from only 15 pairs of flows, and sampling errors must account for a substantial amount of this variability.

Values for station 5 are high and remarkably consistent.

On the basis of Figure 2.2, covariance stationarity will be assumed for time series models of flows for stations 1, 2, 3, 4 and 5.

For stations 6 and 7, low r_j values occur for months 7 and 8, and the highest values are clustered at months 3, 4 and 5. The series for these two stations do not appear to be covariance stationary; if a lag-one model is appropriate for representing these series, (2-22) is to be preferred to (2-21).

Monthly coefficients of skewness C_{sj} are plotted in Figure 2.3. These do not appear to show any particular seasonal pattern; for these data the assumption of a third order stationary model appears reasonable. According to a significance test described by Matalas and Benson (1968), a large number of these coefficients are significantly different to zero at the 95% confidence level.

Skewness coefficients for log series are also plotted in Figure 2.3. By the same confidence test, very few of these are significant. Thus a log-

FIG. 2.2 MONTH-BY-MONTH SERIAL CORRELATIONS

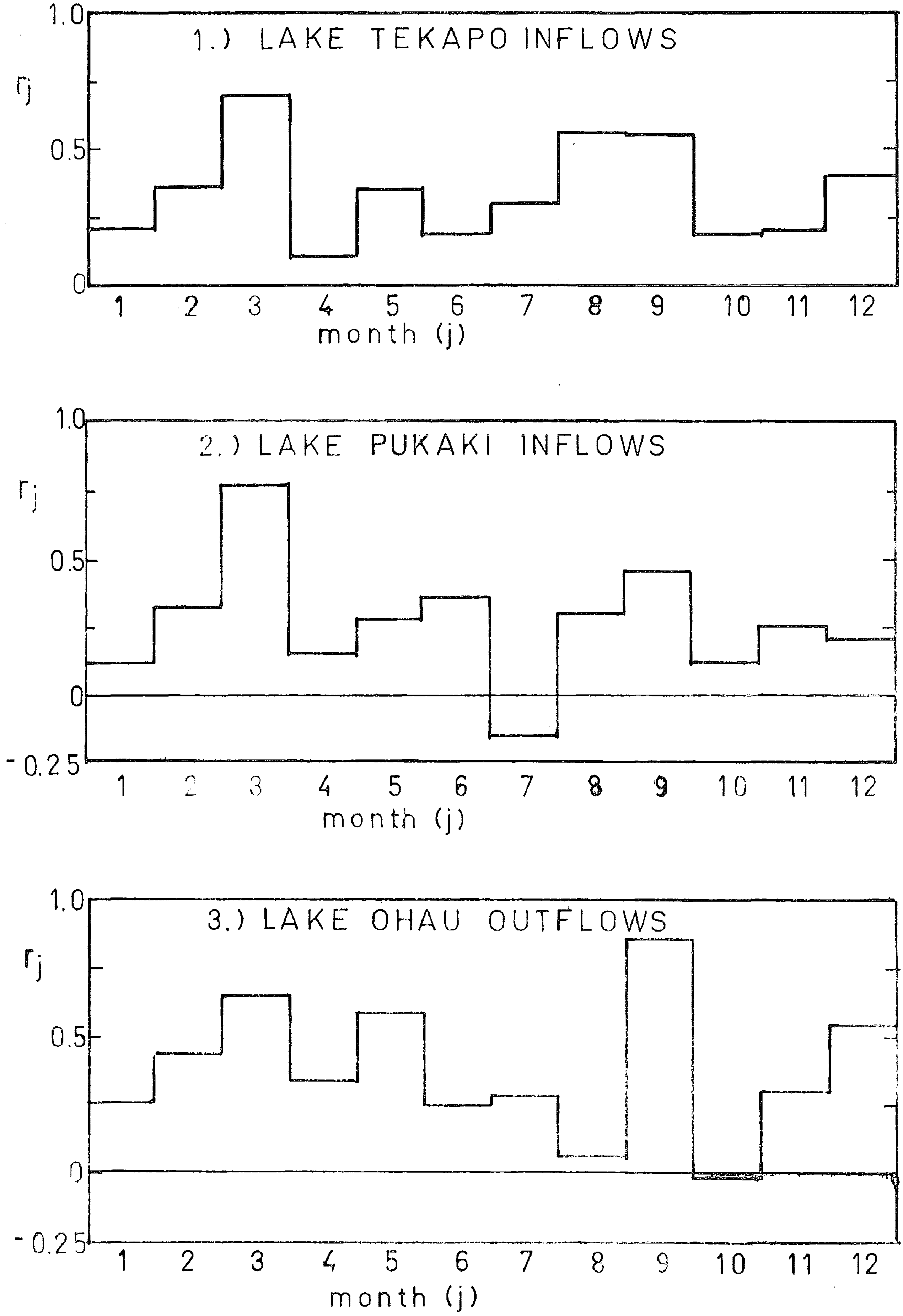


FIG. 2.2(contd)

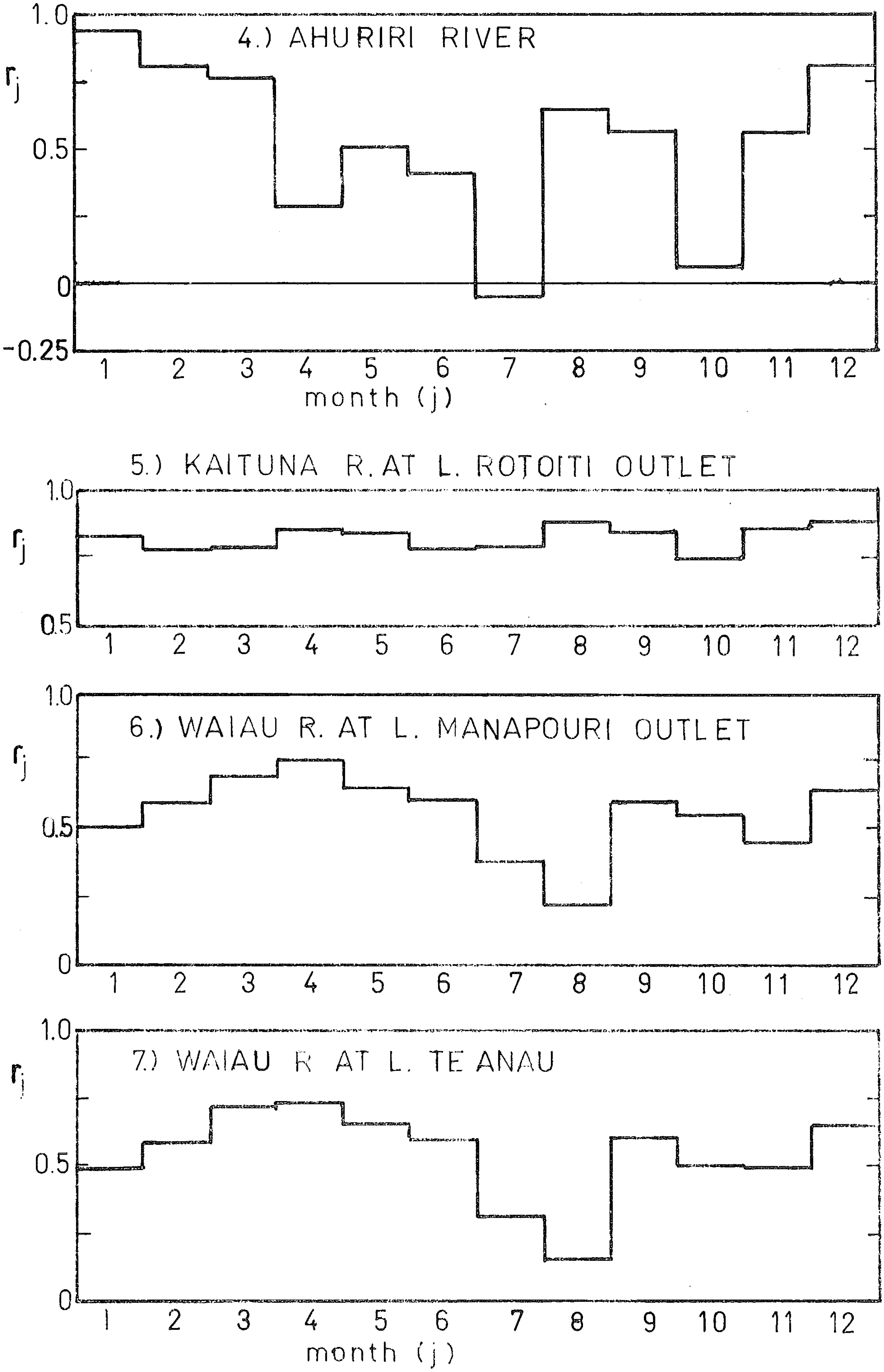

4.) AHURIRI RIVER
1.0
0.5
0
-0.25
r_j
1 2 3 4 5 6 7 8 9 10 11 12
month (j)
5.) KAITUNA R. AT L. ROTOITI OUTLET
1.0
0.5
r_j
1 2 3 4 5 6 7 8 9 10 11 12
6.) WAIAU R. AT L. MANAPOURI OUTLET
1.0
0.5
0
r_j
1 2 3 4 5 6 7 8 9 10 11 12
7.) WAIAU R AT L. TE ANAU
1.0
0.5
0
r_j
1 2 3 4 5 6 7 8 9 10 11 12
month (j)

FIG. 2.3 MONTHLY COEFS OF SKEWNESS

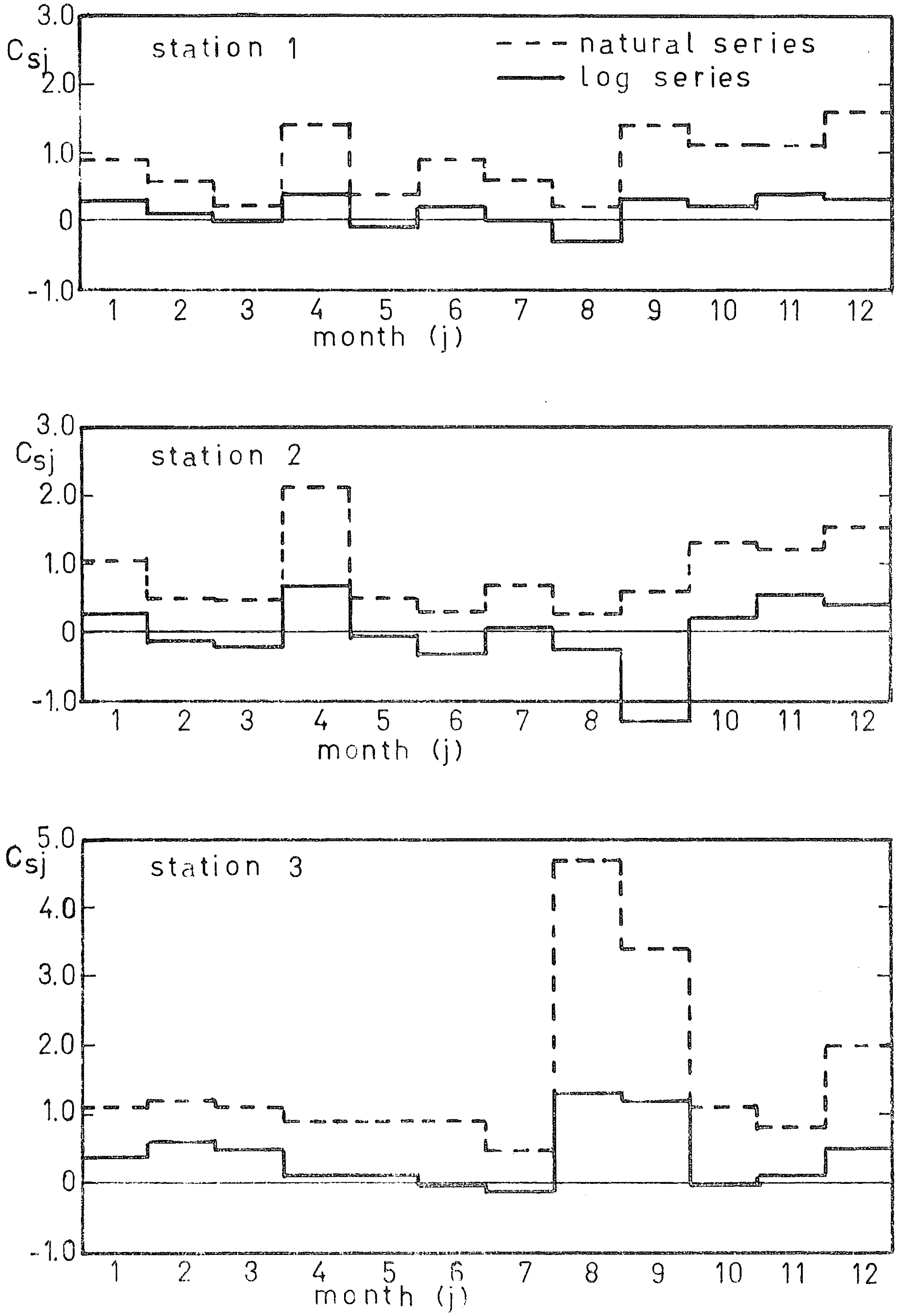
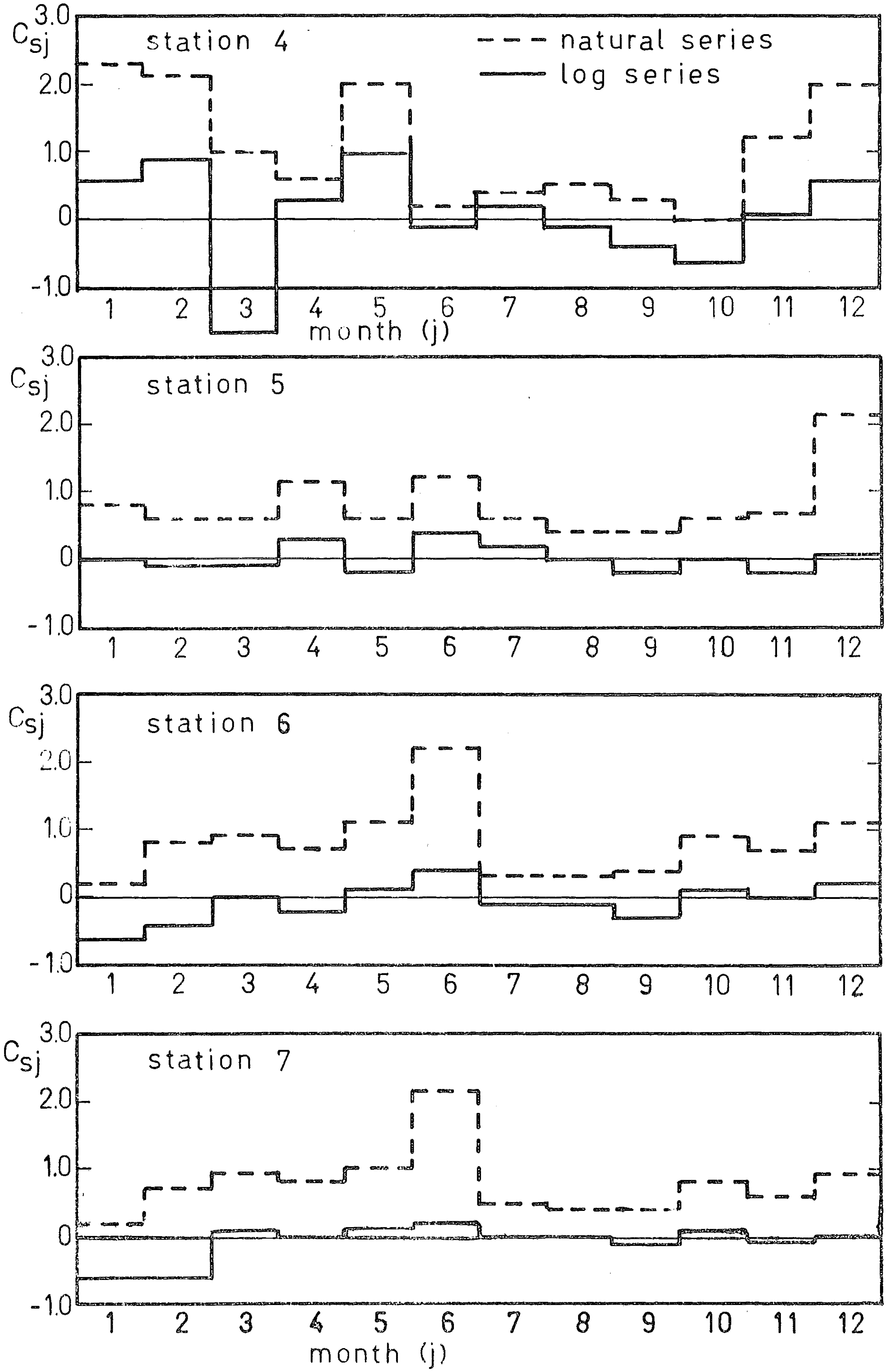


FIG. 2.3(contd)


station 4
natural series
log series
C_{sj}
3.0
2.0
1.0
0
-1.0
1 2 3 4 5 6 7 8 9 10 11 12
month (j)
station 5
station 6
station 7

normal distribution is taken as approximating the distributions of these series, and third order stationarity is imposed.

Further analysis of the data is carried out with $\log_e q_t$ series.

2.5.1 Serial Correlation Analysis

Correlograms calculated for $\log_e q_t$ series are shown in the top graphs of Figures 2.4 - 2.10.

Annual 12 month cycles are very evident for stations 1, 2, 3 and 5. Combinations of annual and 6 month cycles show for series 4, 6 and 7.

Stationary y_t series were obtained by transforming $\log_e q_t$ series with the transform (2-3). Correlograms for these y_t series are the second graphs in Figures 2.4 - 2.10.

The confidence limits given by (2-2) show the first few r(k) values to be different to zero at the 95% confidence level. For larger k values, apparently random variations appear.

Under the assumption of a first order process, Z_t series are obtained using (2-14). Correlograms for these are shown in the third graphs of Figures 2.4 - 2.10.

Very few r(k) values for these correlograms are significant, and these Z_t series might be described as pure random. However, among the r(k) values which do exceed the 95% confidence limit is r(2) for stations 5, 6 and 7. This would suggest that a lag-two model might better fit the y_t series for these stations.

Correlograms for lag-two schemes applied to the y_t series for stations 5, 6 and 7 are the final graphs of Figures 2.8 - 2.10. In these none of the r(1), r(2) or r(3) values are significant.

As a check on the fit of lag-one and lag-two linear autoregression schemes, theoretical variances of Z_t are obtained for the two schemes from (2-8) and (2-11) respectively. These are tabulated with the appropriate calculated variance of Z_t in Tables 2.2 and 2.3. For both lag-one and lag-two schemes, there is very close agreement between theoretical and calculated variances.

FIG. 2.4 CORRELOGRAMS FOR MONTHLY FLOW SERIES FROM STATION 1

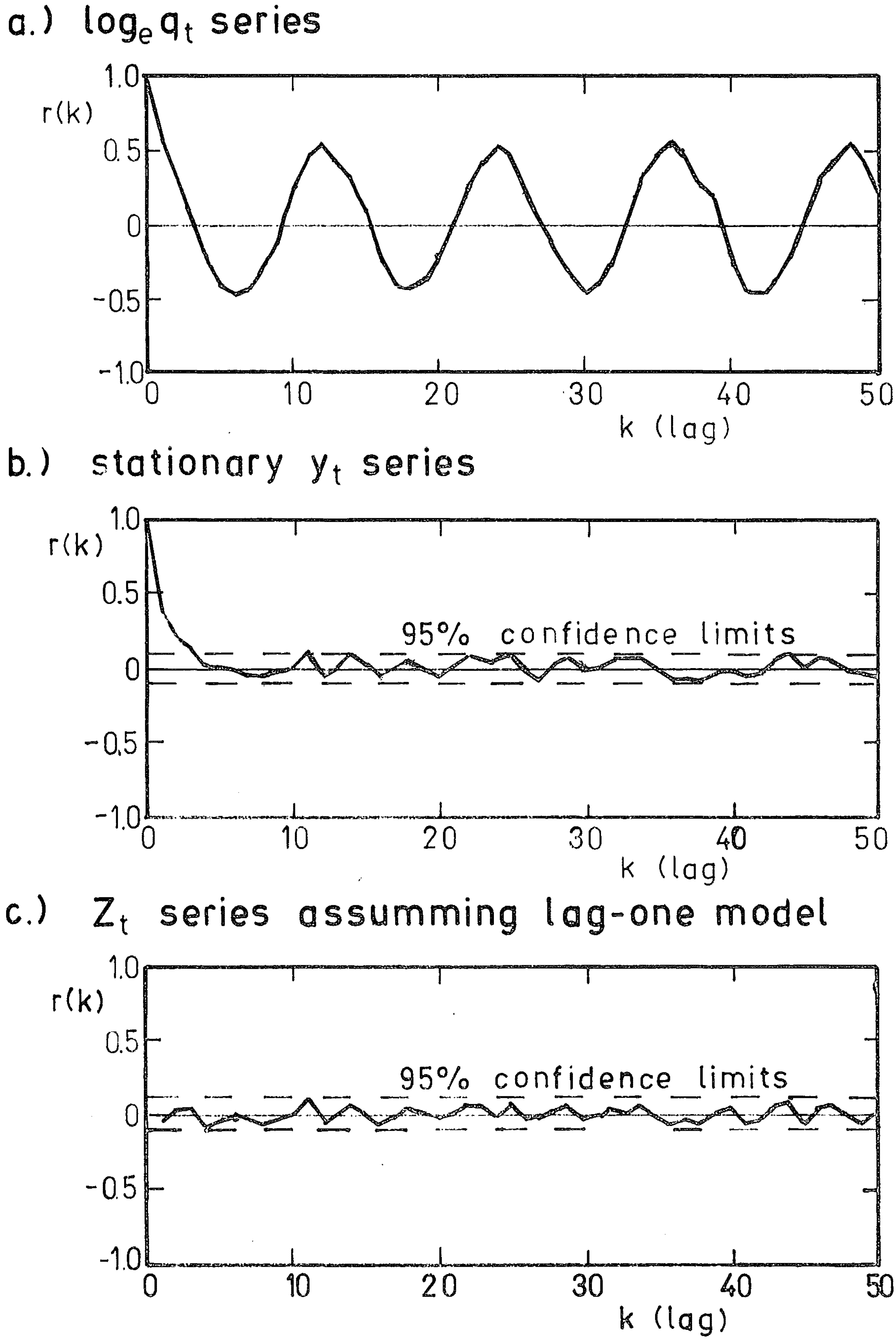


FIG. 2.5 CORRELOGRAMS FOR MONTHLY FLOW SERIES FROM STATION 2

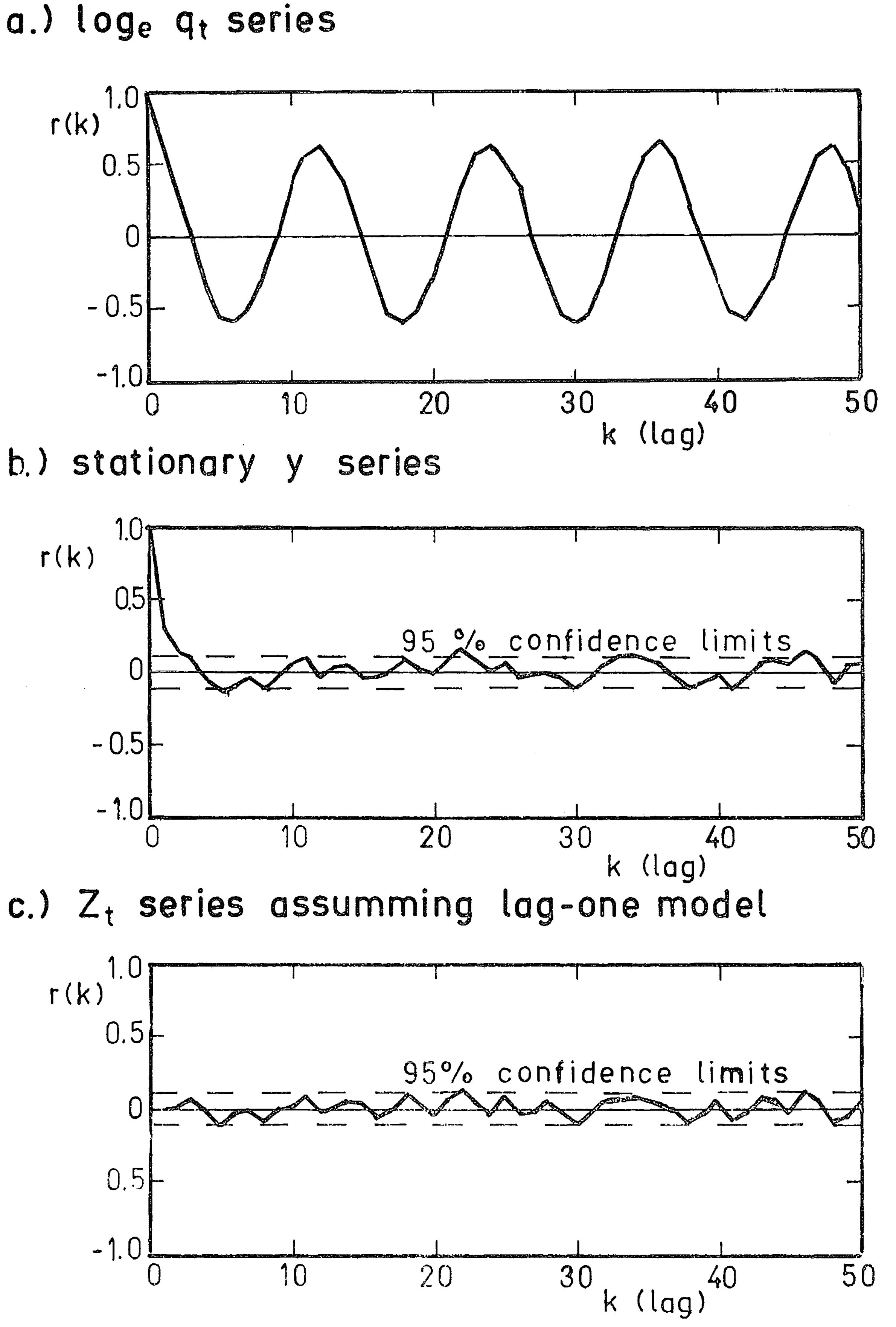
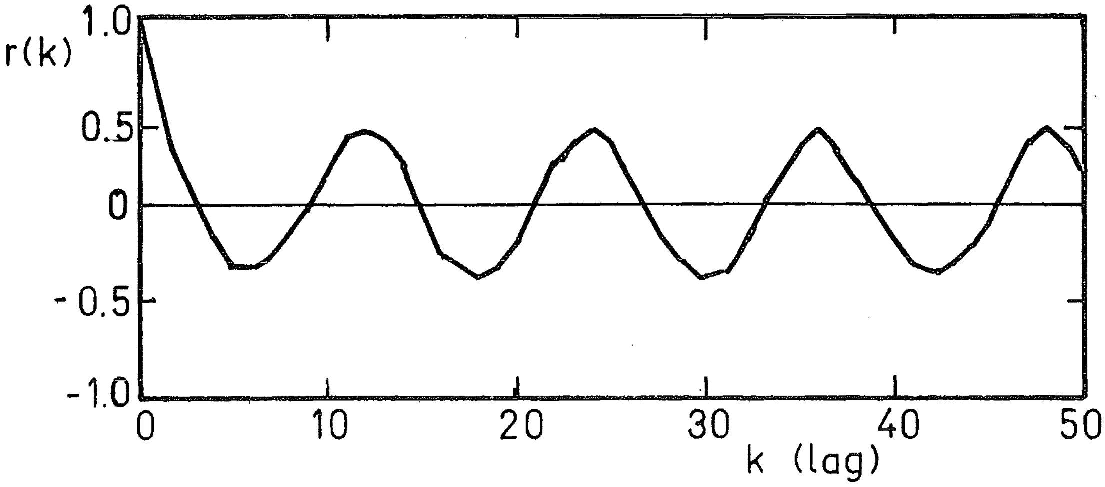
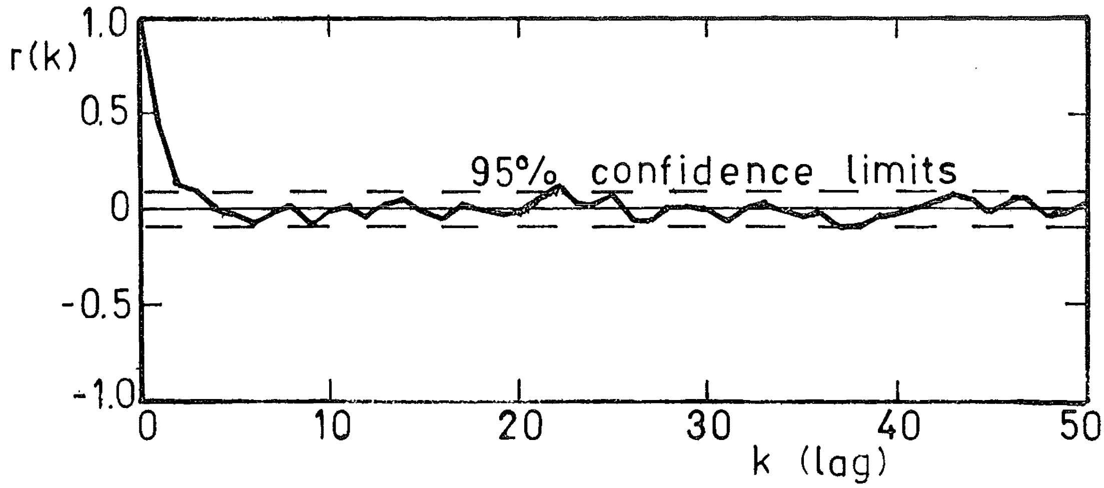


FIG. 2.6 CORRELOGRAMS FOR MONTHLY FLOW SERIES FROM STATION 3

a.) $\log_e q_t$ series



b.) stationary y_t series



c.) Z_t series assumming lag-one model

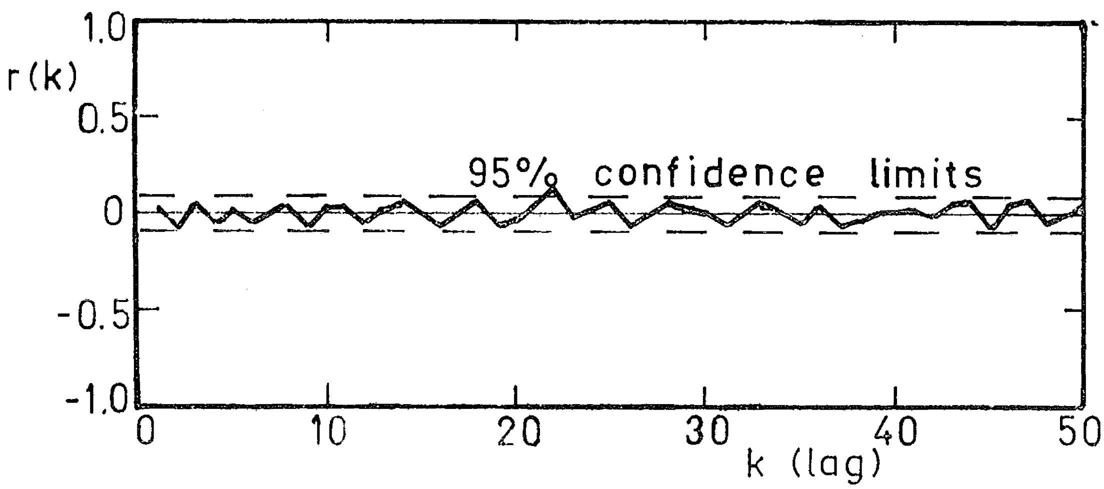
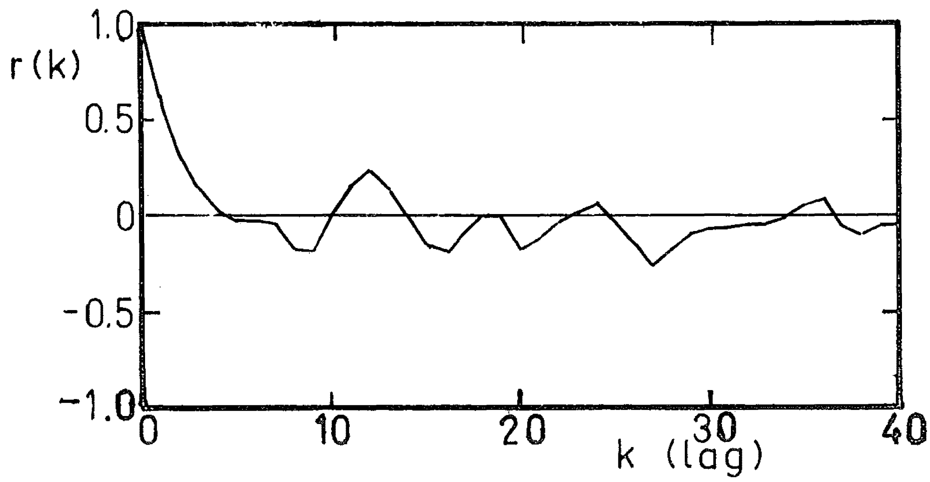
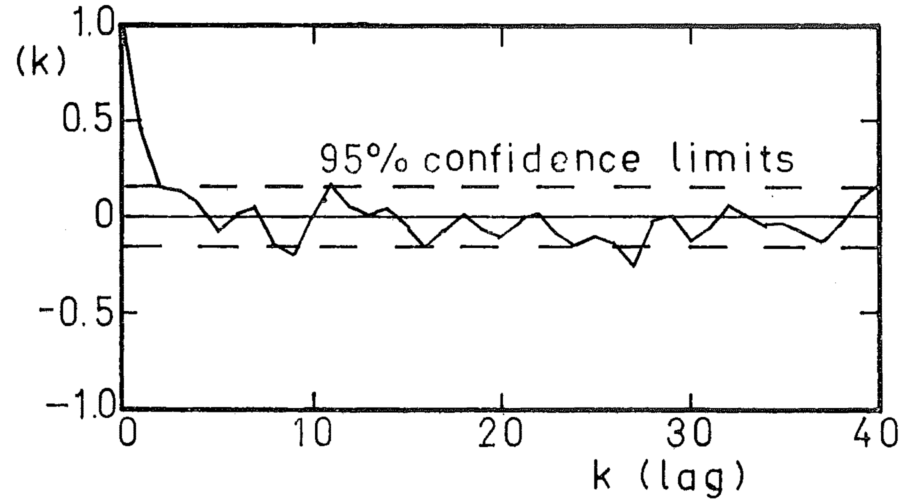


FIG. 2.7 CORRELOGRAMS FOR MONTHLY FLOW SERIES FOR STATION 4

a.) $\log_e q_t$ series



b.) stationary y_t series



c.) Z_t series assuming lag-one model

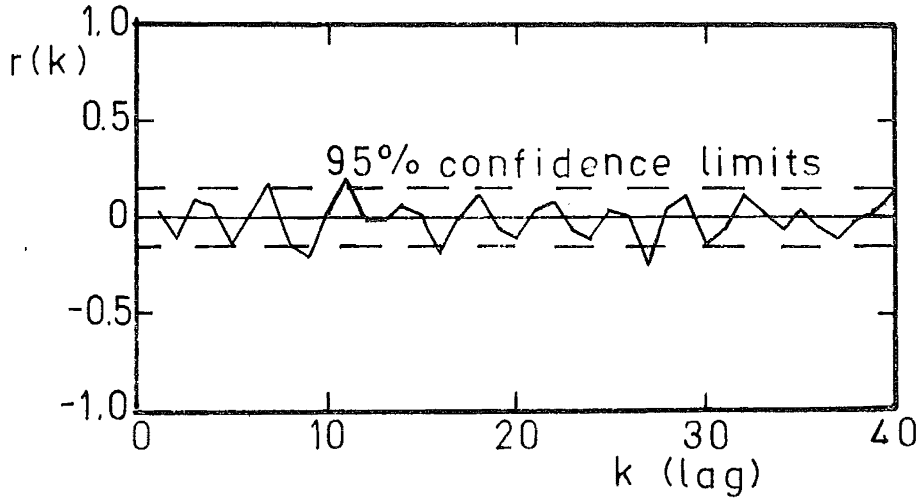


FIG. 2.8 CORRELOGRAMS FOR MONTHLY FLOW SERIES FOR STATION 5

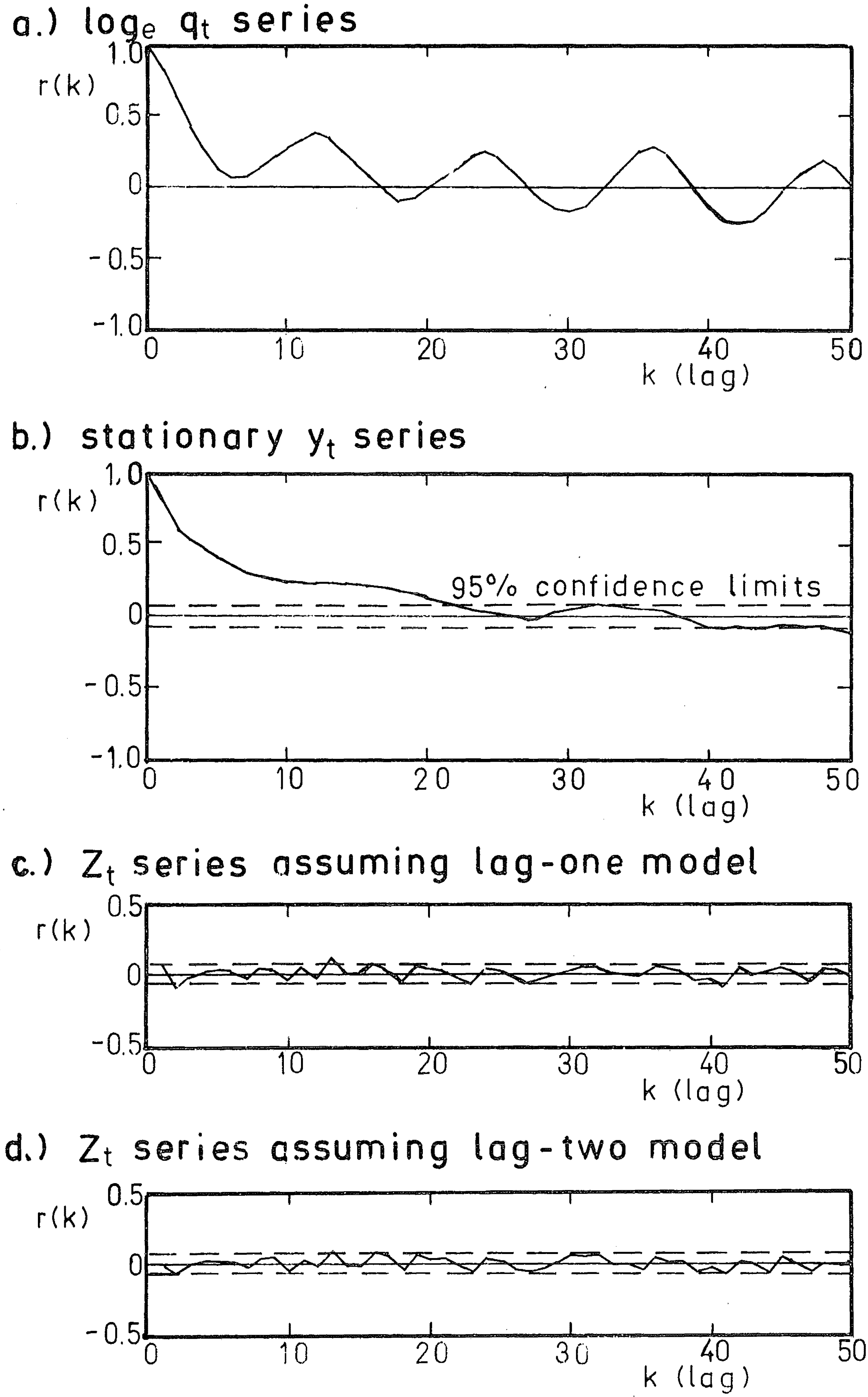


FIG. 2.9 CORRELOGRAMS FOR MONTHLY FLOW SERIES FOR STATION 6

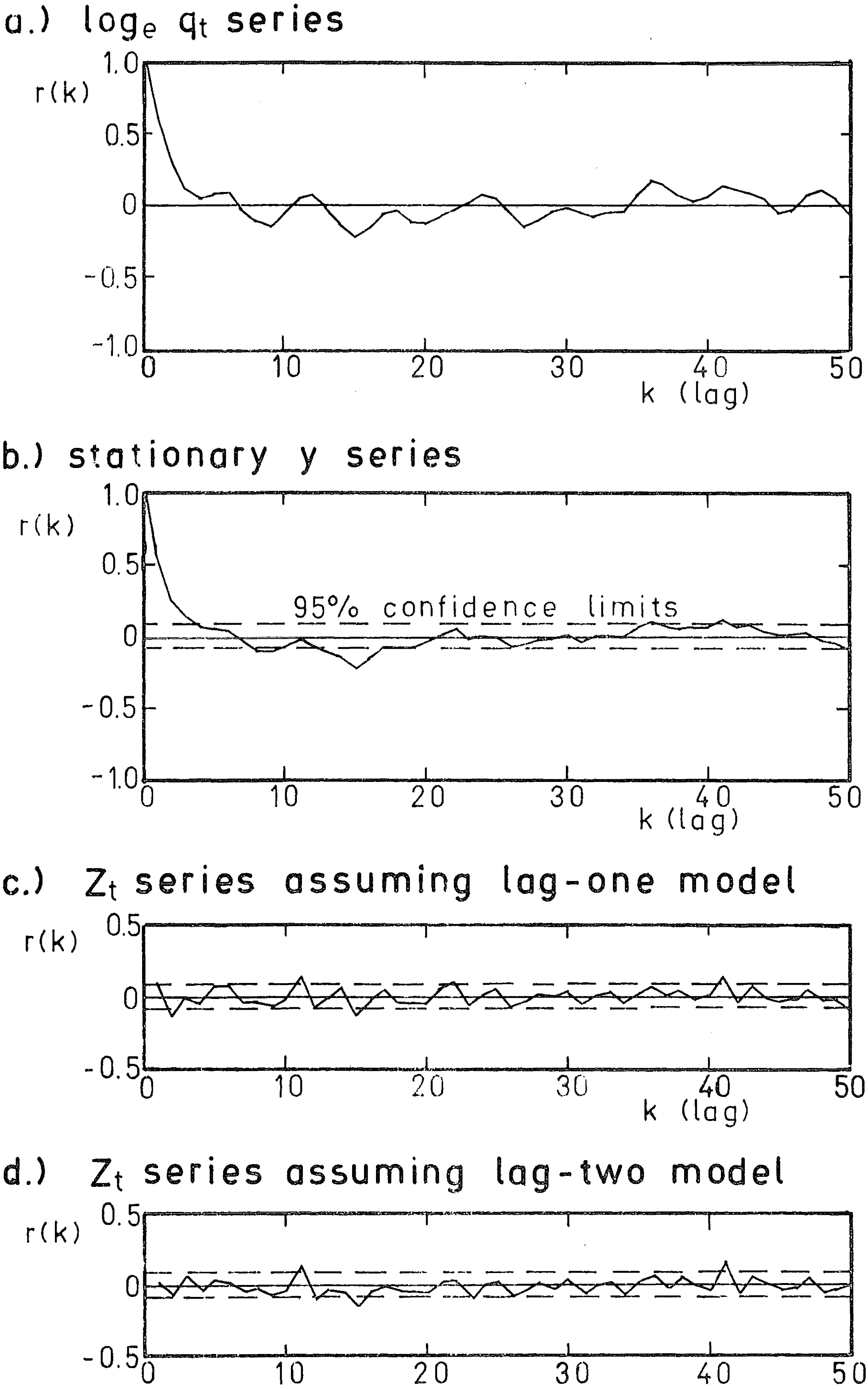


FIG. 2.10 CORRELOGRAMS FOR MONTHLY FLOW SERIES FOR STATION 7

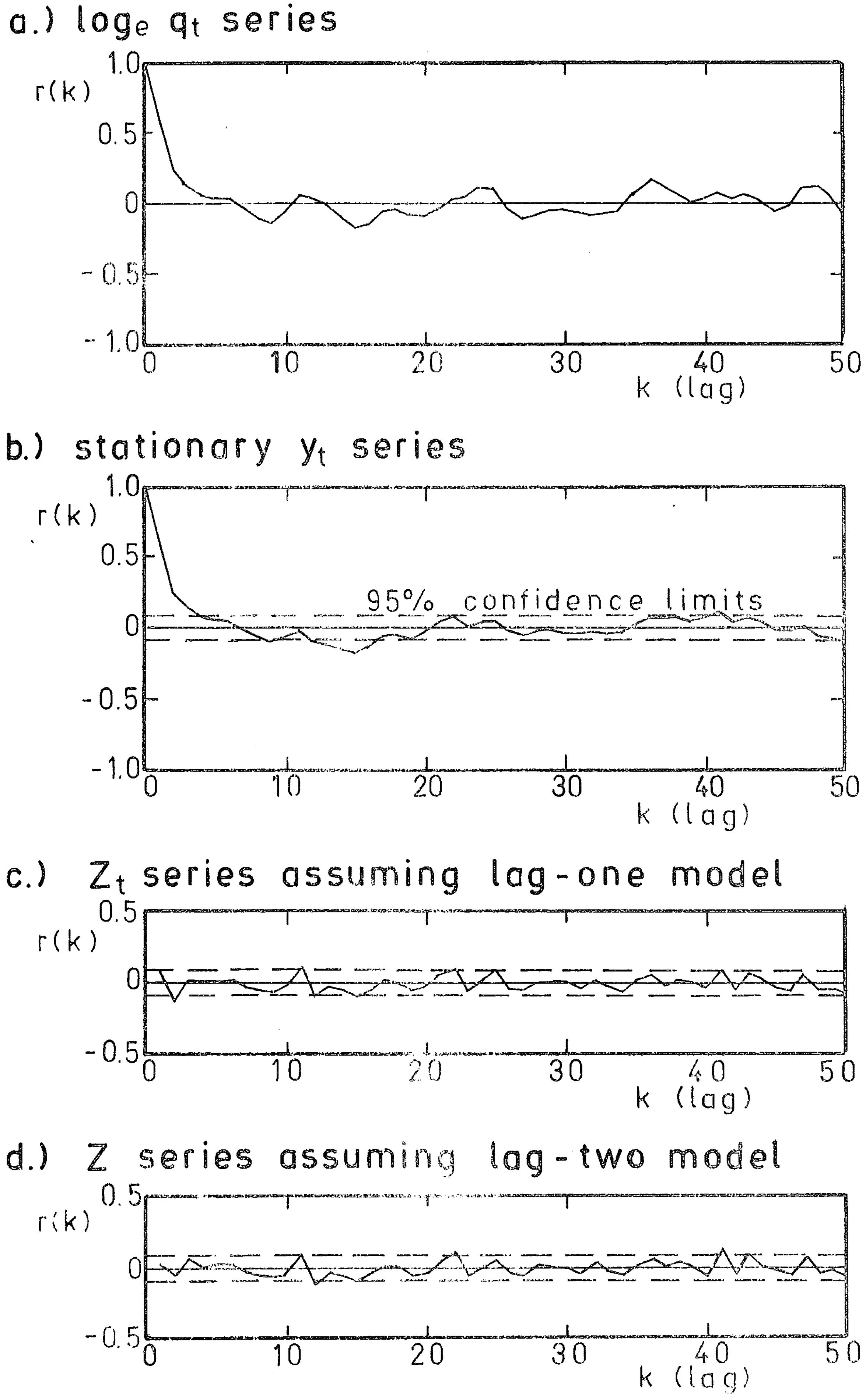


TABLE 2.2

VARIANCE OF Z_t ASSUMING 1ST ORDER AUTOREGRESSION SCHEME

$$y_t = a_1 y_{t-1} + Z_t$$

Station	$a_1 = r_1$	Variance of Z_t series	
		Theoretical	Computed
1	0.379	0.856	0.856
2	0.291	0.915	0.914
3	0.435	0.811	0.811
4	0.451	0.797	0.805
5	0.834	0.304	0.304
6	0.587	0.655	0.655
7	0.566	0.679	0.679

TABLE 2.3

VARIANCE OF Z_t ASSUMING 2ND ORDER AUTOREGRESSION SCHEME

$$y_t = a_1 y_{t-1} + a_2 y_{t-2} + Z_t$$

Station	Correlation coefs.		Autoregression coefs.		Variance of Z_t series	
	r(1)	r(2)	a_1	a_2	Theoretical	Computed
5	.834	.670	.906	-0.086	.302	.302
6	.587	.251	.671	-0.143	.642	.640
7	.566	.241	.633	-0.118	.670	.670

It is concluded that lag-one schemes reasonably well represent monthly flows for all these series.

Covariance stationary models are satisfactory for series from stations 1 to 5, while covariance non-stationary models give a better representation for stations 6 and 7. Also, lag-two models might give a marginal improvement for stations 5, 6 and 7.

Models (2-21) and (2-22) will be used to represent monthly flow series for the seven stations in terms of the means, standard deviations, the skewness, and the serial correlation structure.

2.5.2 Parameter Estimation

Very often water resource designs must proceed without records of reasonable length. In New Zealand, records of more than 10 years length tend to be the exception, rather than the rule.

When developments are proposed for a site with only a brief record available, the question of errors in the estimate of flow parameters arises. In such cases, it is legitimate to ask whether it is worthwhile fitting an autoregression model to the series, or whether the errors in estimate of the parameters are so large as to make the model meaningless.

It may be possible to use rainfall records with a digital rainfall runoff model to estimate additional flows. Alternatively a regression analysis may enable adjacent records and other information to be used to estimate either additional flow data or the important flow parameters for the station. These interesting possibilities are not investigated here, rather the magnitude of errors of estimate which can occur in sample estimates of parameters are examined empirically.

To find the sizes of errors which can result when monthly parameters are estimated from streamflow series of short length, 6 flow records were studied.

With a record of say 40 years length, 5 different samples, each of 8 years length can be formed. The first sample starts with year 1 and ends with year 8, the second year 9 to 16 etc., with the last sample from year 33 to 40. For each of these samples, monthly parameters $\bar{Q}_j$, S_j, C_{sj}, r_j, are estimated and compared with the same parameters for the whole series.

If the parameters from the whole 40 year series are taken as population values which should be estimated, the percentage error for each short record estimate is -

$$e = \frac{(\text{estimated parameter - population parameter})}{\text{population parameter}} \times 100\%$$

For each $\bar{Q}_j$, S_j, C_{sj} and r_j, 5 values for e are obtained, one for each short sample. The average and the maximum of these 5 values are noted. These are averaged over all 12 months, giving an average of average monthly errors, and an average of maximum monthly errors.

In addition, if the short 8 year sample is standardized, a stationary series of 96 values is obtained. The skewness coefficient C_s and the lag-one serial correlation r(1) for this series are evaluated. From the five samples, the average and the maximum errors for these parameters are obtained.

This scheme was applied to the series available. The 40 year record for station 1 gave five independent samples, each 8 years long. In this study, the records for stations 2, 3, 6 and 7 were between 32 and 39 years long, and only four samples were used from each of these; seven samples were drawn from the record for station 5. Data from station 4 is too short to include here.

Average errors of estimate for these six records are given in Table 2.4. Maximum errors are in Table 2.5.

Table 2.4 shows that errors of around 10% are to be expected when average monthly flows $\bar{Q}_j$ are estimated from 8 years of data. Similarly, errors of around 25% are to be expected in estimates of S_j, while errors of more than 100% are frequent in estimates of C_{sj}.

For stations 1, 2 and 3, r_j values are widely variable. For 5, 6 and 7, the expected error is much less. The variability of C_s and r(1) estimates is substantially less.

In general errors tend to increase with the order of the parameter. Estimates of C_{sj} and r_j are quite erratic for a record of eight years length. Rodrígnez-Iturbe (1970) gave similar error analyses for the annual means, variances and 1st order serial correlations for 10 year samples taken from three long flow records. Results similar to those given here were obtained.

TABLE 2.4

AVERAGE ABSOLUTE ERRORS IN PARAMETER ESTIMATES

River			Average absolute error (%)					
	Length Record	No. Samples	Average of 12 monthly values				Standardized Series	
			$\bar{Q}_j$	S_j	C_{sj}	r_j	C_s	r(1)
1	40	5	9	21	150	126	39	20
2	32	4	8	23	171	142	47	33
3	39	4	13	33	735	378	60	5
5	59	7	6	24	102	14	56	6
6	35	4	7	21	116	37	47	16
7	35	4	7	19	107	42	44	17

TABLE 2.5

MAXIMUM ABSOLUTE ERRORS IN PARAMETER ESTIMATES

River		Maximum absolute errors (%)						
	Length Record	No. Samples	Average of 12 monthly values				Standardized Series	
			$\bar{Q}_j$	S_j	C_{sj}	r_j	C_s	r(1)
1	40	5	18	38	305	212	63	52
2	32	4	15	40	362	232	90	76
3	39	4	22	51	1868	638	76	13
5	59	7	15	50	222	36	118	11
6	35	4	12	35	191	73	78	31
7	35	4	13	33	169	84	84	37

Note that these error estimates are all biased downwards. This is because the population estimates of the flow parameters have been obtained from flow records of finite length, and it was from these records that each of the independent short records were drawn.

Nevertheless, there is relevance to synthetic hydrology in these results. Especially if the record is short, there is little point in establishing a flow generating model to preserve parameters which might easily be in error by an order of magnitude. In particular, the skewness coefficient estimates are widely variable. This has several consequences:

(1) It becomes impossible to detect monthly patterns in C_{sj}, thereby dictating the use of a third order stationary model.

(2) Consideration of higher than third order moments, which are subject to evan larger errors of estimate, is impractical.

(3) It becomes difficult to differentiate between alternative non-normal distributions for representing the random series.

Ideally, a design study will incorporate a sensitivity analysis which evaluates the result of errors in these parameter estimates.

2.6 RESULTS FROM DATA GENERATING MODEL

Model (2-21) was used to generate synthetic data for the seven stations. (For the purposes of this study, the series for stations 6 and 7 were assumed to be covariance stationary.)

Normal random deviates were obtained by using the IBM 1130 computing system subroutines GAUSS and RANDU.

To check the computational integrity of the model, $\bar{Q}_j$ and S_j were evaluated for each of 5 - 40 year synthetic sequences. Results for stations 1, 4 and 7 are listed in Tables 2.6, 2.7 and 2.8 with the corresponding historic values. Historic and synthetic estimates of C_s and r(1) are given in Table 2.9. Annual flow statistics are in Table 2.10.

TABLE 2.6

HISTORIC AND SYNTHETIC PARAMETERS FOR STATION 1 (CSD x 1000)

MONTH	MEANS						STD DEVS					
	HISTORIC	SYNTHETIC					HISTORIC	SYNTHETIC				
		1	2	3	4	5		1	2	3	4	5
1	142.4	148.5	136.3	142.3	141.7	138.7	46.6	54.4	40.3	39.4	47.1	48.3
2	126.7	128.7	128.8	120.8	129.4	127.3	46.9	43.8	51.4	42.7	43.7	48.0
3	101.5	100.7	97.9	99.9	102.2	104.3	26.6	29.0	22.3	22.5	24.4	27.3
4	89.2	85.3	88.0	96.4	82.8	85.1	45.6	35.7--	54.5	39.9	41.7	54.5
5	75.8	74.9	80.3	71.0	74.8	66.9--	30.0	32.4	29.7	26.0	35.4	26.0
6	50.2	50.4	52.3	48.8	49.4	49.6	14.0	13.6	16.9	14.1	13.5	14.4
7	46.7	44.7	45.2	44.7	48.9	46.5	13.3	12.8	13.4	13.8	14.1	12.8
8	45.5	43.8	42.2	45.9	47.3	46.1	11.9	15.1	11.9	10.9	13.5	12.1
9	46.3	43.0	45.5	48.5	49.7	51.2	16.7	13.8	16.7	15.8	21.0	19.1
10	89.3	94.8	84.0	85.9	95.1	84.6	37.3	39.5	31.8	28.8--	46.1	32.3
11	106.1	108.1	105.2	114.1	107.8	101.9	35.3	33.5	33.0	31.7	36.1	40.1
12	140.1	150.0	148.6	144.6	136.7	131.1--	40.6	42.7	52.3++	39.1	37.6	28.4--

TABLE 2.7

HISTORIC AND SYNTHETIC PARAMETERS FOR STATION 4 (CSD x 1000)

MONTH	MEANS						STD DEVS					
	HISTORIC	SYNTHETIC 1	2	3	4	5	HISTORIC	SYNTHETIC 1	2	3	4	5
1	33.2	34.0	30.3	28.5--	34.0	31.4	20.5	18.3	13.7--	13.1--	18.9	16.0--
2	28.3	26.8	26.2	26.0	29.3	27.8	20.0	18.1	11.5--	16.0--	17.8	18.8
3	24.7	25.1	26.3	23.8	25.2	25.1	10.6	13.8++	11.6	10.5	14.2++	14.8++
4	23.6	25.7	21.9	22.7	21.2	24.6	10.3	9.9	7.6--	13.2	8.4	9.7
5	34.4	30.5	34.0	31.7	36.2	37.0	16.4	13.9	18.8	12.5--	18.2	15.4
6	28.0	26.7	27.7	28.4	28.2	27.5	7.3	7.7	7.4	8.0	6.2	6.8
7	26.1	25.2	28.4	25.7	27.5	27.4	8.2	7.6	9.7	6.8	8.1	7.7
8	26.1	26.7	28.2	27.2	29.7	26.6	9.2	8.4	10.2	9.3	11.5	9.6
9	30.8	32.6	34.4	32.7	31.3	31.6	11.3	11.3	14.5	14.3	10.9	10.8
10	41.4	42.3	39.8	41.2	42.7	40.0	11.2	12.3	10.3	10.1	12.1	11.8
11	42.9	46.5	42.2	40.7	41.0	42.0	16.3	16.5	13.9	17.7	11.9--	14.5
12	46.4	48.6	43.9	41.2--	44.7	48.2	23.3	23.6	15.8--	15.9--	21.2	20.0

TABLE 2.8

HISTORIC AND SYNTHETIC PARAMETERS FOR STATION 7 (CSD x 1000)

MONTH	MEANS						STD DEVS					
	HISTORIC	SYNTHETIC					HISTORIC	SYNTHETIC				
		1	2	3	4	5		1	2	3	4	5
1	318.4	319.3	321.5	308.5	320.2	295.2	89.6	98.0	92.4	85.0	112.0	86.7
2	284.5	287.7	289.5	271.7	283.5	278.0	109.9	113.0	120.5	113.9	122.0	124.1
3	289.3	282.3	271.5	270.8	300.1	284.1	108.2	96.8	104.7	102.3	112.2	106.0
4	280.6	278.2	249.3	276.9	297.7	285.2	103.6	129.5	100.8	81.4--	121.1	105.2
5	311.0	306.3	294.5	328.3	333.1	330.9	119.9	108.1	131.6	117.1	149.1	128.8
6	256.7	270.2	245.8	261.0	259.2	254.4	77.7	85.9	66.7	72.5	85.9	66.8
7	252.9	256.0	254.7	274.6	250.0	246.6	66.0	69.6	60.9	70.5	59.3	83.5
8	232.5	242.7	247.7	236.0	223.9	224.2	61.5	70.9	67.7	64.4	54.5	59.8--
9	243.8	251.4	241.8	241.9	239.2	239.4	66.9	75.6	61.7	61.2	60.1	62.0
10	347.4	357.7	348.3	330.2	356.3	346.6	107.4	110.3	115.7	100.4	81.2--	106.2
11	339.1	334.8	328.4	327.0	356.7	330.8	99.2	108.8	78.8--	97.0	80.5--	94.2
12	350.4	338.4	345.4	356.5	352.2	335.2	106.6	76.3--	116.0	101.0	98.4	125.4

TABLE 2.9

SKEWNESS AND SERIAL CORRELATION COEFFICIENTS FOR HISTORIC AND SYNTHETIC SERIES, STATIONS 1, 4 AND 7

Station Number	Skewness Coefficient C_s						Serial Correlation Coefficient r(1)					
	Historic Series	Synthetic series					Historic Series	Synthetic series				
		1	2	3	4	5		1	2	3	4	5
1	.88	.72	.76	.62	.71	.78	.34	.37	.40	.37	.39	.38
4	.97	1.12	.74	.86	.87	.86	.52	.44	.45	.43	.43	.46
7	.79	.76	.83	.72	.72	.83	.55	.58	.58	.56	.55	.52

TABLE 2.10

ANNUAL FLOW STATISTICS FOR HISTORIC AND SYNTHETIC SERIES, STATIONS 1, 4 AND 7.

Station Number	Mean annual flows C.S.D. x 1000						Annual Standard Deviations C.S.D. x 1000					
	Historic Series	Synthetic series					Historic Series	Synthetic series				
		1	2	3	4	5		1	2	3	4	5
1	1060	1073	1054	1062	1067	1036	173	160	155	112	170	153
4	387	395	386	373	392	390	73	67	62	62	72	64
7	3507	3536	3500	3486	3583	3459	520	681	594	588	620	589

Station Number	Skewness coefficient of annual flows						Serial correlation coef. of annual flows					
	Historic Series	Synthetic series					Historic Series	Synthetic series				
		1	2	3	4	5		1	2	3	4	5
1	.27	.34	.94	1.06	.54	.69	.06	-.05	-.04	.03	-.02	-.04
4	1.43	.98	.94	.37	.92	1.52	.21	.26	.14	.13	.19	.22
3	.82	.79	1.24	.85	.74	.67	-.05	-.02	.09	-.02	.01	.35

Standard errors of estimate for each synthetic $\bar{Q}_j$ and S_j are calculated as $\frac{S_j}{\sqrt{40}}$ and $\frac{S_j}{\sqrt{80}}$ respectively. Approximate 95% confidence limits of $\pm$ twice the standard error were established. If synthetic $\bar{Q}_j$ and S_j are not significantly different from their historic counterparts, the historic values should lie within this band at least 95% of the time.

In Tables 2.6, 2.7 and 2.8, those synthetic values which are significantly greater than the historic values are marked ++, while those significantly less are marked --. In all, only 84% of the synthetic standard deviations are within the range of $\pm$ 2 standard errors. This is less than expected if the differences between historic and synthetic values arise by chance.

Three possible causes for this lack of resemblance between historic and synthetic values are:

(1) The method of fitting the log-normal distribution.

(2) The deviates from the random number generator not having zero mean and unit variance.

(3) Estimates of means and standard deviations of variables following log-normal distributions are only approximately normally distributed.

To check the first, synthetic data was generated with the log-normal distribution fitted by the method of moments and using the same sequence of random numbers. This resulted in 87% of the standard deviations lying within the 95% range. Thus, despite the theoretical advantages of the maximum likelihood estimates, the method of moments is preferred here, although the difference is not large.

The second cause was checked by generating again the sequences of random numbers used to obtain the synthetic data. After setting them into the monthly groups as used by the flow generating process, these numbers were analysed to check for conformity to the requirements of zero mean and unit variance. This analysis proved the random number generator; the generated numbers could not be distinguished from random deviates having zero mean and unit variance.

The third point is taken up in Section 3.4.4 of Chapter 3.

With the confidence test used here and a log-normal distribution fitted by the method of moments, reasonable but not entirely satisfactory agreement between historic and synthetic parameters has been achieved. The test described is rigorous; it can be used to show that other synthesized results, for instance those reported by Thomas and Fiering (1962), are too variable.

An alternative less rigorous way to show correspondence between historic and synthetic sequences is illustrated by Harms and Campbell (1967). Synthetic $\bar{Q}_j$ and S_j values were estimated for 25 different sequences. For each monthly class, the means $\bar{x}_Q$ and $\bar{x}_S$ and the standard deviations σ_Q and σ_S were obtained. The ranges $\bar{x}_Q \pm 2\sigma_Q$ and $\bar{x}_S \pm 2\sigma_S$ were plotted for comparison with the historic (population) values. In this way, agreement between historic and synthetic sequences was illustrated.

Similar results for stations 1, 4 and 7, but with estimates from 5, instead of 25 sequences, are illustrated in Figures 2.11 to 2.13. These indicate good agreement between historic $\bar{Q}_j$ and S_j and synthetic values obtained from 5-40 year samples.

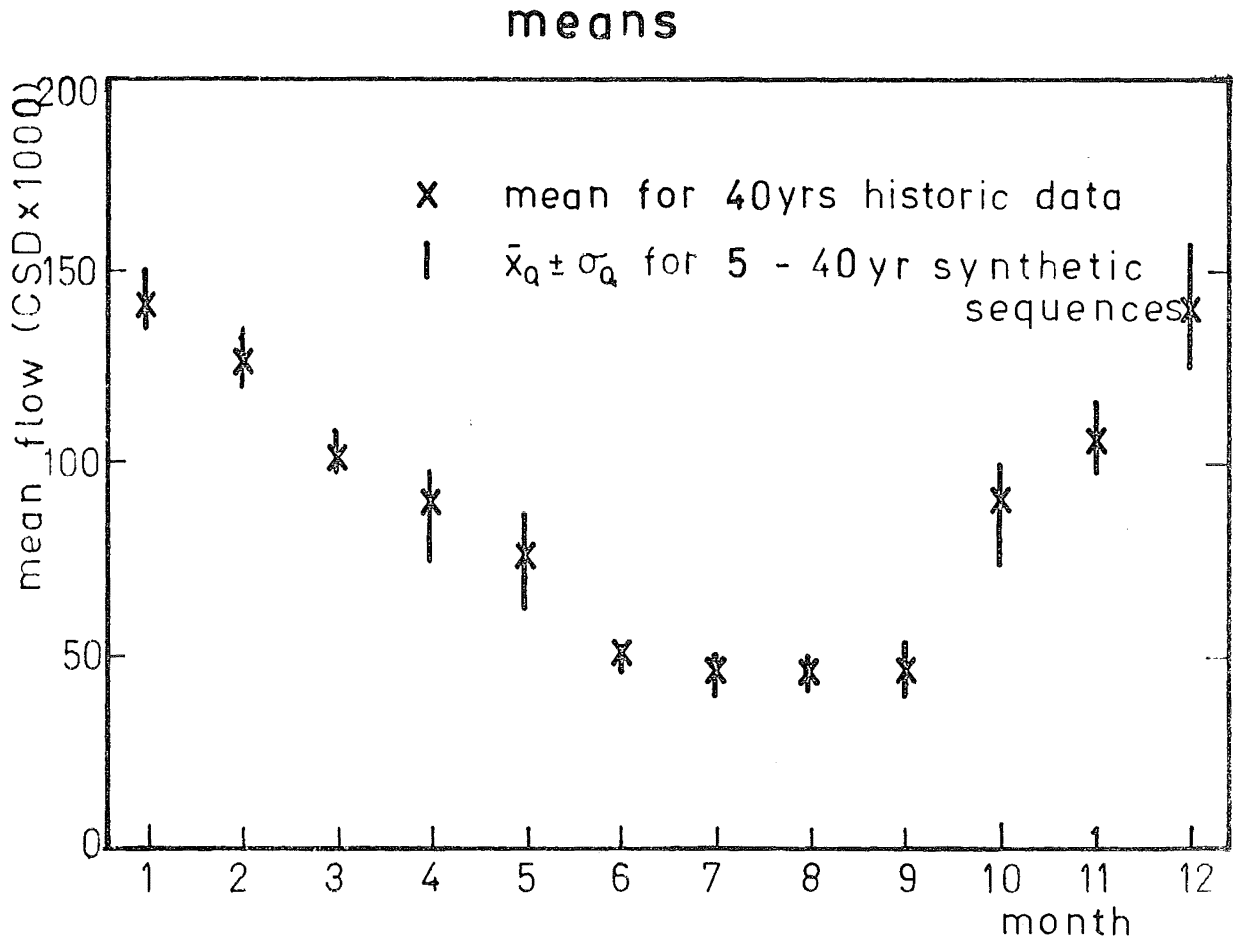
On the basis of these results, and those of Tables 2.9 and 2.10, the computational integrity of the model is demonstrated.

2.7 SUMMARY

Time series models for representing monthly streamflow series have been described. Flow series from seven stations were analysed and in every case good representation of the flow logarithms was achieved by a Markov lag-one model. This linear autoregression model represents the series in terms of the monthly means, variances, serial correlation and skewness.

The model can be used to generate synthetic sets of data which resemble the historic data in terms of the monthly flow parameters. It is observed that annual flow statistics are preserved by this procedure.

FIG. 2.11 HISTORIC & SYNTHETIC MEANS & STANDARD DEVIATIONS FOR STATION 1



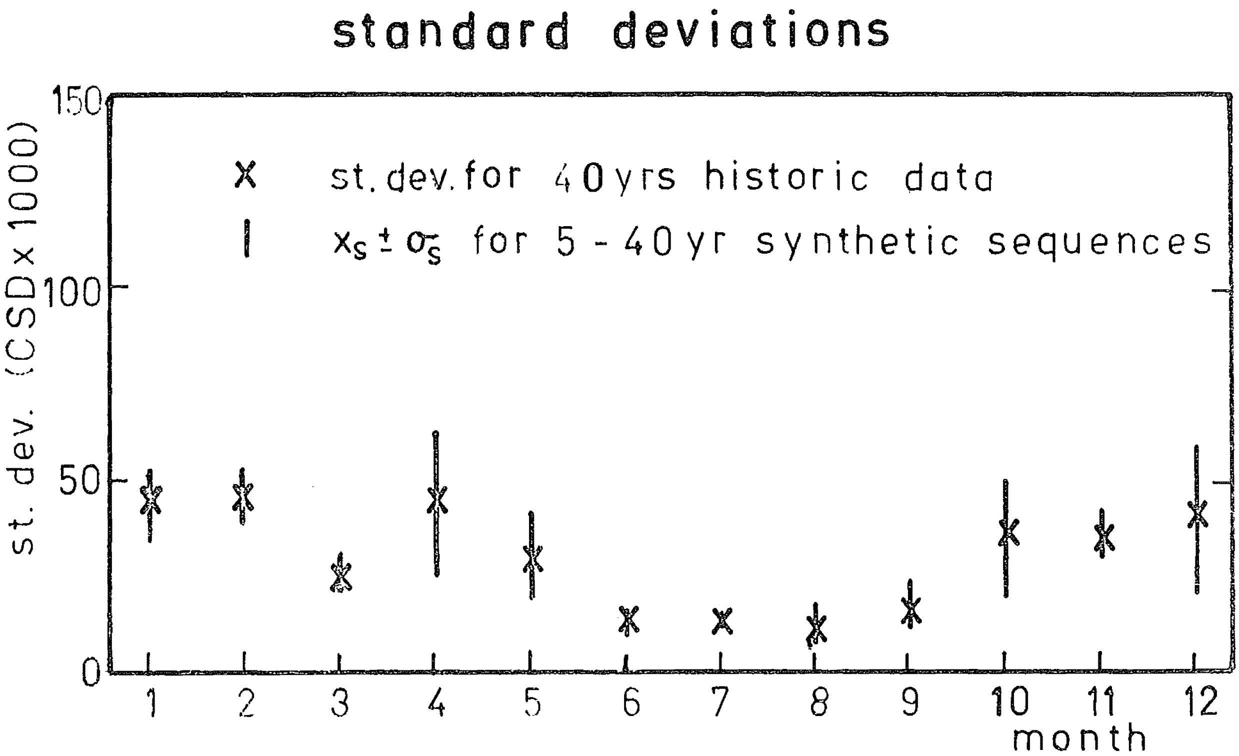
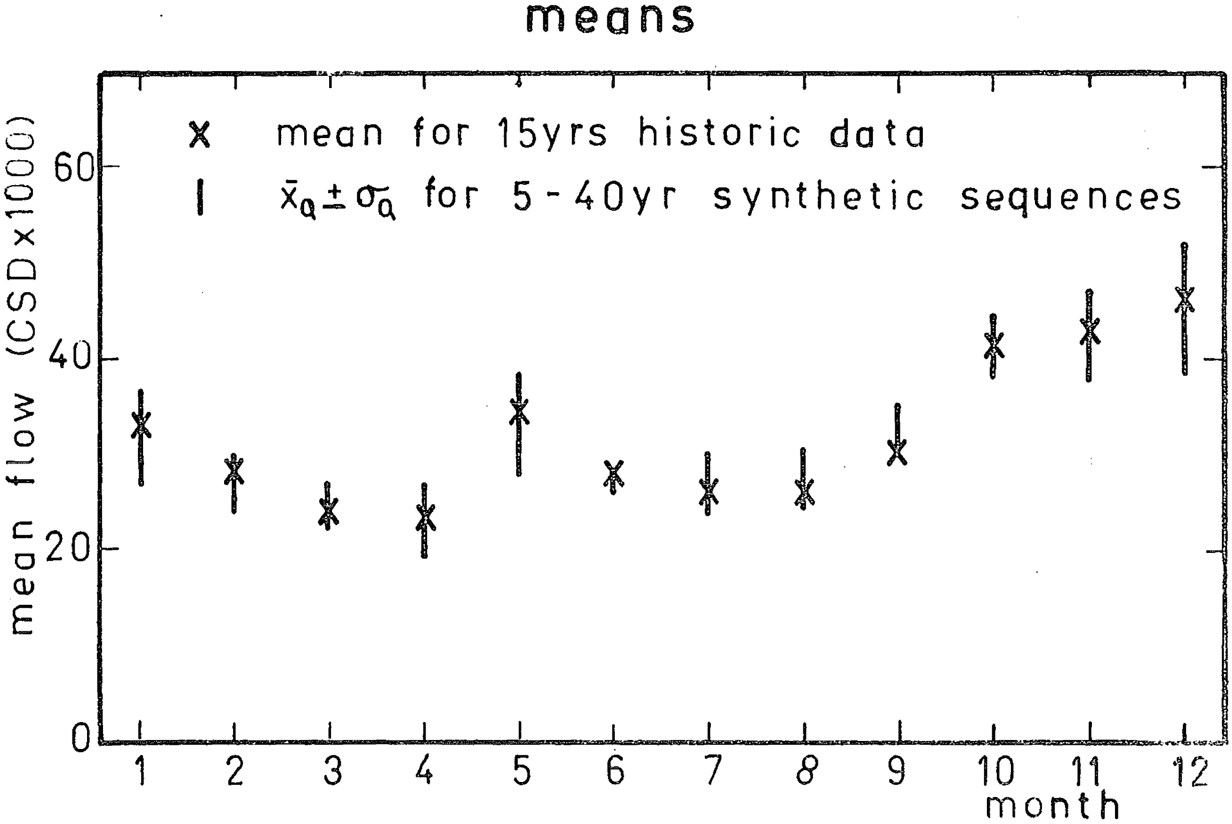


FIG. 2.12 HISTORIC & SYNTHETIC MEANS & STANDARD DEVIATIONS FOR STATION 4



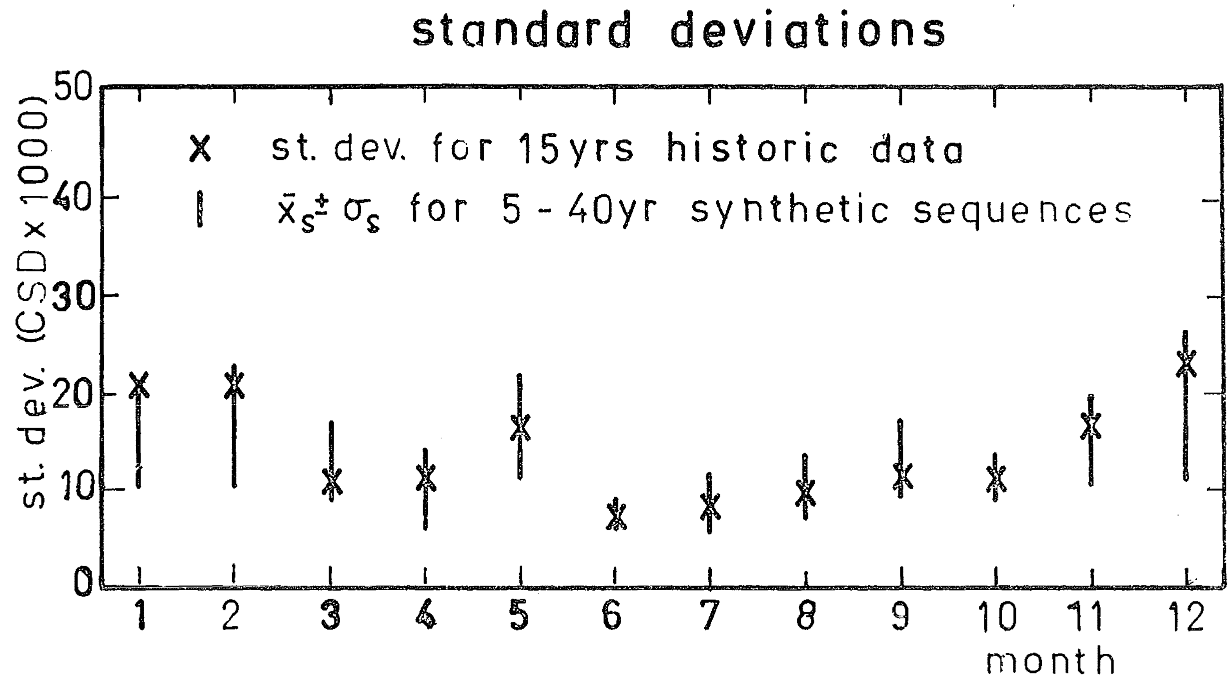
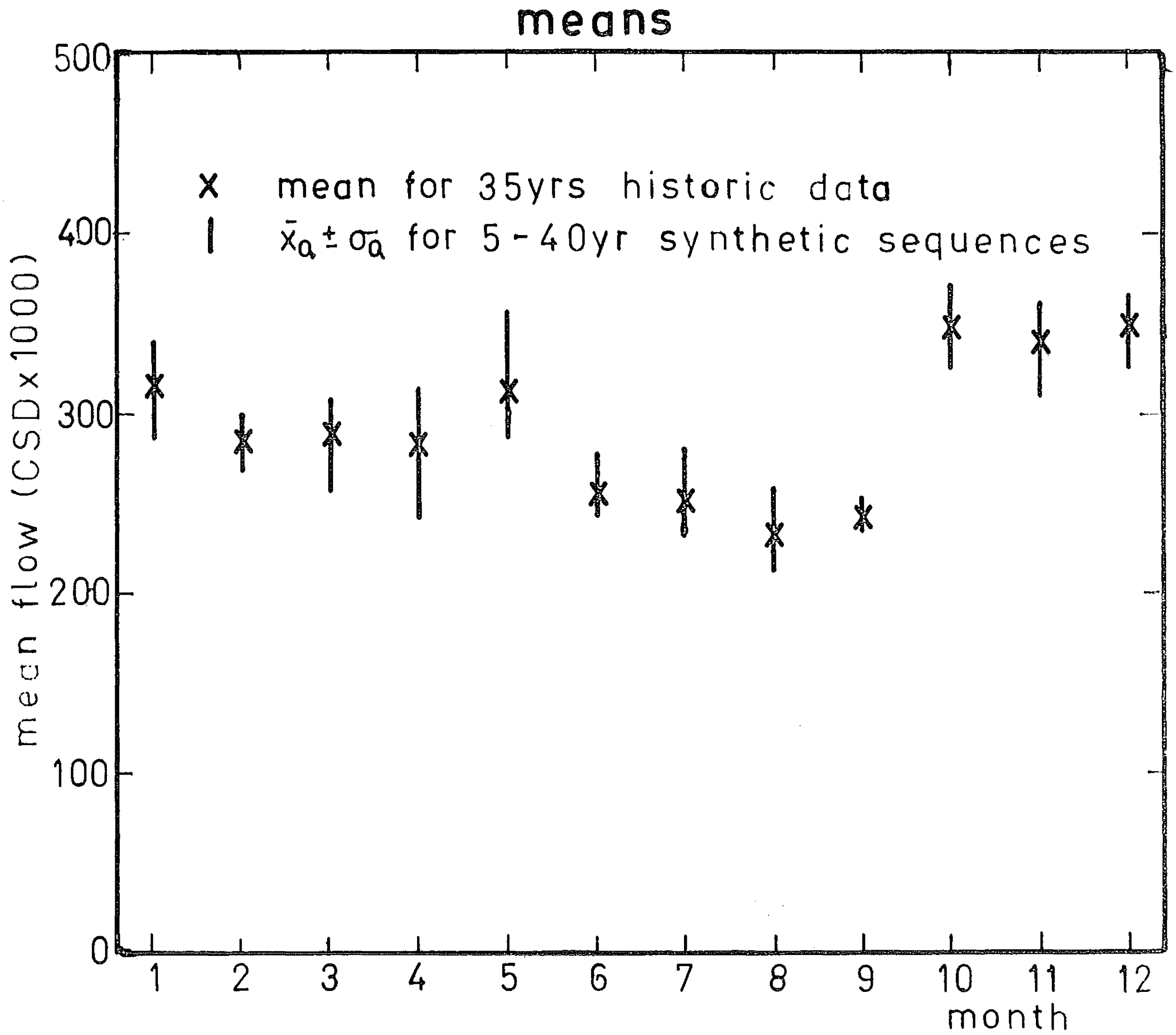
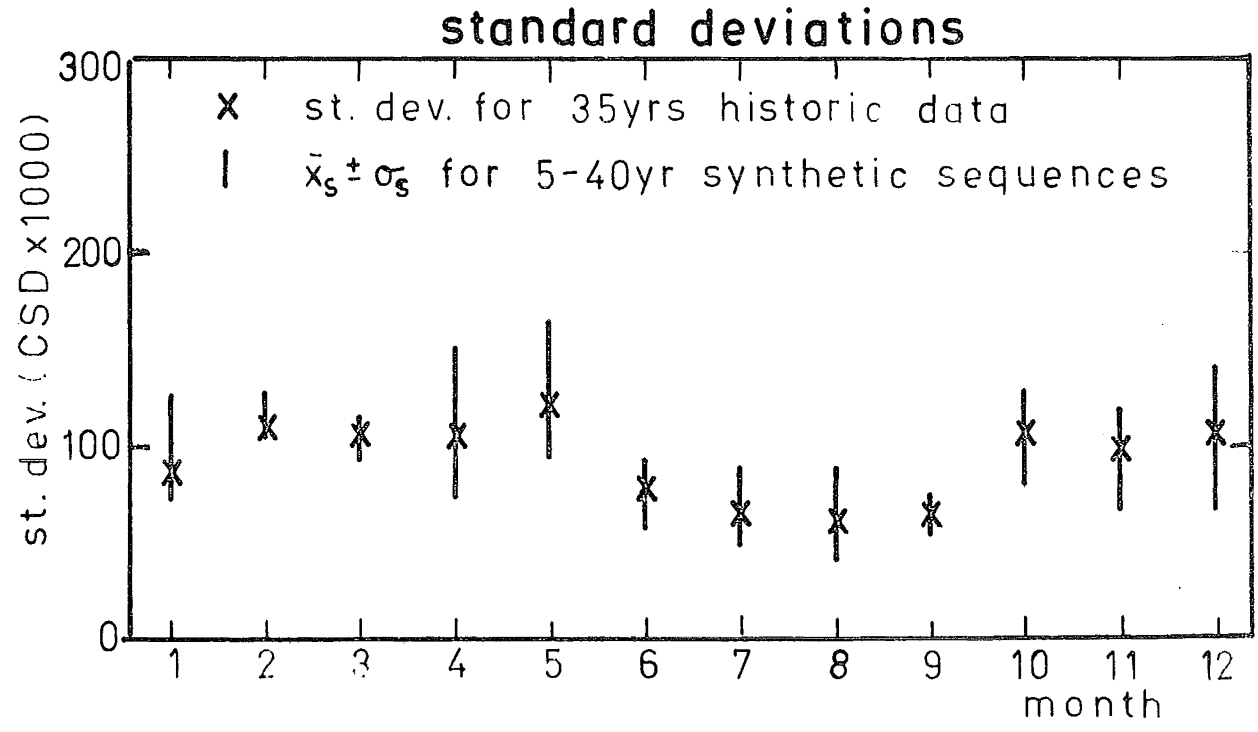


FIG. 2.13 HISTORIC & SYNTHETIC MEANS & STANDARD DEVIATIONS FOR STATION 7





Thus the generated data can be described as synthetic flows. These synthetic flow series, which represent possible combinations of future flows at a gauging station, are useful in water resource system design studies.

CHAPTER THREE

STREAMFLOW ANALYSIS AND SYNTHESIS II, THE MULTI STATION MODEL

3.1 INTRODUCTION

Many design studies are concerned with systems which encompass one or more river basins. In such systems, inflows can occur at several different localities.

A flow generating model for this situation must generate data series for several flow stations. The single station flow models described in Chapter 2 are inadequate for this; they do not take account of the interdependencies of flows which can exist within a region.

This chapter describes a multistream generating model which preserves the parameters of flows at several different stations, the persistence relationship of flows through time, and the interstation relationships. The model is quite general. Provided a covariance stationary lag-one Markov process is appropriate for the flow series for different stations, and the stationary series for each station are either normally or log-normally distributed, the model can be used to generate weekly, monthly or annual interrelated flow series for any computationally reasonable number of stations.

To illustrate the method, data series for stations 1, 2, 3 and 4 will be used. These stations are all within the Waitaki River basin. In Chapter 5 a study to find optimum operating rules for the hydro-electric power system within this basin is described. This study will use the multistream model to simulate sequences of inflows into the system.

3.2 INTERSTATION CORRELATIONS

The interdependencies between flow series from different stations are measured by the lag-zero cross correlation coefficients between stationary series. These coefficients have been estimated between the monthly y_t series for stations 1, 2, 3 and 4. As indicated in Table 2.1, the y_t series are not all of the same length, and the estimates are based on the parts of the different y_t series which occur together. These estimates in matrix form, are given by M_o,

$$M_o = \begin{bmatrix} 1.00 & & & \\ 0.87 & 1.00 & & \\ 0.75 & 0.79 & 1.00 & \\ 0.71 & 0.67 & 0.80 & 1.00 \end{bmatrix}$$

Here, the i, jth element of M_o is the best estimate of the lag-zero cross correlation between series for stations i and j. Because M_o is symmetric, the upper elements are not shown. M_o may be described as the "lag-zero cross correlation matrix."

The time dependent features of the series have already been examined in Chapter 2. For the Waitaki basin, the interstation dependencies appear to be much more important than the time dependencies.

3.3 LITERATURE REVIEW

Several generating schemes are available which model the relationship between stations, as well as the individual station parameters.

3.3.1 Thomas-Fiering Model

Thomas and Fiering (1962) describe a generating method which requires the selection of a "key" station. For this key station a synthetic series is generated using a single station model, (2-21) or (2-22). For each of the other satellite stations, a record is then generated using a bivariate linear regression equation. This equation is very similar to (2-21) but has the satellite statistics q_{t-1}, $\bar{Q}_{j-1}$, S_{j-1} replaced by q_t, $\bar{Q}_j$, S_j for the key station, and $r(1)$ is replaced by the observed lag-zero cross correlation coefficient between the satellite and the key station. Thus -

$$q_t^s - \bar{Q}_j^s = r^{sk} \frac{S_j^s}{S_j^k} \cdot (q_j^k - \bar{Q}_j^k) + t_t \cdot S_j^s \cdot \left\{1 - (r(1)^{sk})^2\right\}^{\frac{1}{2}}$$

The superscripts s and k refer to values for the satellite and key station respectively.

This technique preserves the relationship between the key station and the different satellites, but not the correlation between satellites, nor the serial correlation at satellite stations. Efforts to avoid these limitations, and also the rather subjective selection of a "key" station have led to several alternative formulations.

3.3.2 Beard Model

Beard (1965) used a set of multiple linear regression relationships to generate interrelated data series for flow stations. This method used an approximation to a gamma distribution to represent the normalized logarithms of monthly flow volumes. The approximation to the gamma distribution was Pearson type III. Skewness was not used directly in the synthetic process. Instead, the synthesis was done in terms of normal deviates of logarithms, which were subsequently converted to skewed deviates of logarithms by an empirical relationship.

3.3.3 Fiering Model

A multivariate streamflow synthesis procedure was proposed by Fiering (1964). This scheme used principal component analysis to transform a matrix of standardized observations at m real stations for N time periods into n vectors each of length t, with $n \leqslant m$. The n vectors were in fact principal components of the original data, successive components explaining progressively smaller amounts of variance.

These components were mutually independent of each other and represented n hypothetical stations, all independent of one another. Synthetic data series, with the same statistics as these components were then generated with the single stream model. Finally, an inverse linear transform was applied to transform the synthetic data for the n hypothetical stations into synthetic data for m real stations.

This method had a weakness in that the original transform to get the principal components assumed that successive elements of data in the principal components were serially independent. In fact this assumption was not valid. The result was that expected values of the serial correlation coefficients for the synthetic series for the real stations were not the respective historic values, but a value resulting from a combination of the historic coefficients (Matalas, 1967b).

This limitation only becomes serious when the serial correlation coefficient differs substantially for different stations. Since most studies which use interrelated synthetic data are likely to be based on one basin, or several adjacent basins with similar climatic regimes, wide variations in the serial correlation coefficients would be hard to justify, and one value for all stations might be quite reasonable.

For stations 1, 2, 3 and 4, r(1) values ranged between 0.26 and 0.52, with an average value of 0.36. In this situation it might be reasonable to preserve the average value with this technique.

3.3.4 Matalas Model

An alternative multivariate scheme which preserves both the serial and cross correlation coefficients is given by Matalas (1967b). It is defined by the relationship -

$$Y_t = A\, Y_{t-1} - B\, \epsilon_t \qquad (3\text{-}1)$$

where Y_t and Y_{t-1} are defined as vectors of standardized observations at m stations at times t and (t-1) respectively. Thus y_t^p is the standardized observation at the pth station at time t.

ϵ_t is a vector of length m of standardized independent random variates.

A and B are (m x m) matrices whose elements are defined so that the synthetic sequences resemble the historic sequences in terms of the lag-zero cross correlations and the lag-one serial correlations.

Effectively, (3-1) is the multistream analogue of the single stream Markov lag-one model given by (2-7). Its implementation for generating inter-related synthetic data is similar to the single station scheme.

With the initial values $Y_o = 0$, synthetic $Y_1, Y_2, Y_3 \ldots\ldots Y_t$ values can be obtained from the recursive relationship (3-1). Finally the inverse linear transform

$$q_t^p = S_j^p \; y_t^p + \bar{Q}_j^p \tag{3-2}$$

is applied to obtain the non-stationary q_t^p series, the superscript p referring to the pth station. If a transform was made by taking $\log_e q_t^p$, the natural values are obtained by exponentiation.

Two assumptions implied in this formulation are:

(1) Each of the m series making up the historic data is third order and covariance stationary.

(2) All the y_t^p series belong to the same population. In this study they are all taken as log-normally distributed.

For the data from stations 1, 2, 3 and 4, these assumptions are not restrictive.

It remains to define A and B in (3-1). Post-multiply (3-1) by Y_{t-1}^T, the (1 x m) matrix obtained by transposing the vector Y_{t-1},

$$Y_t \; Y_{t-1}^T = A \; Y_{t-1} \; Y_{t-1}^T + B \, \epsilon_t \; Y_{t-1}^T$$

Taking expected values, and noting that $E\,(B \; \epsilon_t \; Y_{t-1}^T) = 0$, since ϵ_t is defined as an independent array and $E\,(\epsilon_t) = 0$;

$$E\,(Y_t \; Y_{t-1}^T) = E\,(A \; Y_{t-1} \; Y_{t-1}^T)$$

$$= A \; E\,(Y_{t-1} \; Y_{t-1}^T) \,.$$

Let $M_1 = E(Y_t \cdot Y_{t-1}^T)$ and $M_o = E(Y_{t-1} \cdot Y_{t-1}^T)$,

thus

$$M_1 = A \cdot M_o \qquad (3\text{-}3)$$

M_o is a (m x m)matrix with diagonal elements $E(y_t^p \cdot y_t^{p^T}) = 1$ since y_t^p has unit variance. Similarly, off diagonal elements are $E(y_t^p \cdot y_t^q) = \rho^{(p)(q)}(0)$, the lag-zero cross correlation between stations p and q, $1 \leqslant p \leqslant m$, $1 \leqslant q \leqslant m$ and $p \neq q$.

M_1 is a(m x m)matrix with diagonal elements,

$E(y_t^p \cdot y_{t-1}^p) = \rho^{(p)}(1)$, the lag-one serial correlation at p, $1 \leqslant p \leqslant m$. The off diagonal elements,

$E(y_t^p \cdot y_{t-1}^q) = \rho^{(p)(q)}(1)$, are the lag-one cross correlations between p and q, $1 \leqslant p \leqslant m$, $1 \leqslant q \leqslant m$, $p \neq q$.

Similarly, post-multiplying (3-1) by Y_t^T,

$$Y_t \cdot Y_t^T = A\, Y_{t-1} \cdot Y_t^T + B \cdot \epsilon_t\, Y_t^T.$$

Taking expected values,

$$E(Y_t \cdot Y_t^T) = A \cdot E(Y_{t-1} \cdot Y_t^T) + B \cdot E(\epsilon_t \cdot (Y_{t-1}^T \cdot A^T + \epsilon_t^T \cdot B^T))$$

thus,

$$M_o = A \cdot M_1^T + B \cdot B^T \qquad (3\text{-}4)$$

Equations (3-3) and (3-4) are solved for A and B. (From (3-3),

$$A = M_1 \cdot M_o^{-1} \qquad (3\text{-}5)$$

Substitute for A in (3-4),

$$B \cdot B^T = M_o - M_1 M_o^{-1} \cdot M_1^T \qquad (3\text{-}6)$$

Equation (3-6) does not give B directly, but the product $B \cdot B^T$. To obtain B, principal component analysis can be used, (Matalas, 1967b), but because the product $B \cdot B^T$ is symmetric, a direct recursive solution is also possible (Young and Pisano, 1968).

3.4 IMPLEMENTATION OF MATALAS MODEL

3.4.1 Previous Result

Young and Pisano (1968) list results obtained from a computer program developed to implement this technique. 21 years of historic data from eight sites were used to illustrate its application.

Parameters for one group of synthetic series 100 years long were listed for comparison with historic parameters. If the technique is correct and functioning properly, synthetic parameter estimates might be expected to deviate from the historic values by amounts dependent on their respective errors of estimate. Thus about 95% of the synthetic parameters might be expected to differ from their respective population values by less than two standard deviations.

In fact, in the example given by Young and Pisano, many more than 5%, about 35% of the synthetic standard deviations differed by more than two standard errors from the population values (Jennings, 1969). In addition, most of the synthetic lag-one serial correlation coefficients were less than the population values, and many of the synthetic skewness coefficients were greater than the population values.

This lack of resemblance between historic and synthetic results may be due to:

(1) an apparent failure to transpose the final M_1 matrix in (3-6),

(2) use of the maximum likelihood method instead of the method of moments in estimating parameters for log-normal distributions.

A further limitation of this example was that it required all the historic data to cover the same time span. This results in a wastage of information if the historic records cover unequal time periods and only segments of some records can be used.

3.4.2 Estimation and Consistency of Correlation Matrices

Matrices M_0 and M_1 contain least squares estimates of lag-zero and lag-one correlations respectively.

Lag-one cross correlations, given by $\rho^{(p)\,(q)}(1)$, $p \neq q$, are not of special interest; to facilitate their computation, Matalas (1967b) suggests that they be calculated as -

$$\rho^{(p)\,(q)}(1) = \rho^{(p)\,(q)}(0)\,.\,\rho^{(p)}(1) \tag{3-7}$$

for all p and q, $p \neq q$.

To make more efficient use of the longer historic records, $\rho^{(p)\,(q)}(0)$ can be estimated from the parts of the records from stations p and q which cover the same time period. The correlation matrix so obtained must be consistent, which means that correlations estimated between different pairs of records must be compatible. Fiering (1968) has investigated methods for handling inconsistent matrices. In particular, he has shown that for a correlation matrix to be consistent, it is necessary and sufficient to show that the matrix is non-negative definite. This condition is realised when all the eigen values of the correlation matrix are non-negative.

3.4.3 Application to Waitaki Data

The method summarized by (3-1) was programmed for use within a computer. Features of the program and the computing facilities used are described in Appendix I.

Data series for the stations 1, 2, 3 and 4, all within the Waitaki River Basin were used as input for this study. The algorithm is further used in Chapter 5 to generate flow data for these stations for use as input in a simulation study. The program is quite general however, and the number of stations included is restricted only by the computational facilities available. Output from the program consists of synthetic data of the required length and the parameters for the historic and synthetic data and the correlation matrices.

Correlation matrices estimated for the standardized Y_t series, without taking logarithms were;

$$M_o = \begin{bmatrix} 1.00 & & & \\ 0.87 & 1.00 & & \\ 0.75 & 0.79 & 1.00 & \\ 0.71 & 0.67 & 0.82 & 1.00 \end{bmatrix}$$

$$M_1 = \begin{bmatrix} 0.34 & 0.30 & 0.26 & 0.24 \\ .23 & .26 & .21 & .17 \\ .28 & .30 & .33 & .30 \\ .37 & .35 & .42 & .52 \end{bmatrix}$$

Correlation matrices for the y_t series obtained by standardizing the $\log_e q_t$ series were

$$M_o = \begin{bmatrix} 1.00 & & & \\ 0.85 & 1.00 & & \\ 0.79 & 0.79 & 1.00 & \\ 0.71 & 0.63 & 0.81 & 1.00 \end{bmatrix}$$

$$M_1 = \begin{bmatrix} 0.38 & 0.32 & 0.30 & 0.27 \\ .25 & .29 & .23 & .19 \\ .34 & .34 & .43 & .35 \\ .32 & .28 & .36 & .45 \end{bmatrix}$$

Eigen values for both the M_o matrices given here are positive; therefore the matrices are consistent.

Time series analyses of the Waitaki data has been reported in Chapter 2. The generation of interrelated data series reported here is carried out in terms of logarithmic series.

Matrices A and B obtained by solving (3-5) and (3-6) were -

$$A = \begin{bmatrix} .376 & 0 & 0 & 0 \\ 0 & .291 & 0 & 0 \\ 0 & 0 & .427 & 0 \\ 0 & 0 & 0 & .448 \end{bmatrix}$$

$$\text{and} \quad B = \begin{bmatrix} .927 & 0 & 0 & 0 \\ .814 & .502 & 0 & 0 \\ .715 & .226 & .506 & 0 \\ .638 & .063 & .365 & .505 \end{bmatrix}$$

In obtaining B, the recursive solution described by Young and Pisano (1968), was used.

3.4.4 Verification of Results

With these results 5 - 40 year sets of interrelated data were generated for the four stations. Synthetic means and standard deviations are tabulated in Tables 3.1 - 3.5 with the corresponding historic values. Historic and synthetic estimates of C_s and r(1) are given in Table 3.6. Lag-zero cross correlations are given in Table 3.7, and annual flow parameters are in Table 3.8.

Agreement between historic and synthetic means, correlations and skewness is good, as evidenced by Tables 3.1 to 3.7. Synthetic standard deviations do not conform to the historic values as closely as expected. Only about 75% of the standard deviations, instead of the expected 95% differ by less than two standard errors from the historic values.

In terms of the resemblance between historic and synthetic parameters, these results are an improvement over those described by Young and Pisano (1968). Nevertheless the divergence between historic and synthetic standard deviations has not been satisfactorily accounted for.

In part, at least, this divergence is connected with the use of a log-normal distribution. That this is so is shown by using the same sequences of random numbers to generate synthetic data series, but assuming that the

TABLE 3.1

SUMMARY OF SYNTHETIC MEANS AND ST DEVS FOR 40 YEARS DATA,

DATA SET 1

	SITE NUMBER							
	1		2		3		4	
MONTH	MEAN	ST DEV	MEAN	ST DEV	MEAN	ST DEV	MEAN	ST DEV
1	147.2	54.7	260.4	86.5	127.4	53.9	39.4	27.5 ++
2	109.9 --	36.1 --	224.6 --	86.5	92.7	38.8	27.3	16.2 --
3	95.3	29.0	184.0 --	54.9	76.6	31.9	23.3	12.4
4	68.7 --	29.9 --	120.6 --	58.8 --	64.6 --	27.0 --	21.6 --	11.0
5	75.0	26.5	103.7	43.7	69.7	30.6	33.5	16.8
6	53.5	15.2	60.4	17.8	50.8	15.9	29.8	8.1
7	47.7	13.5	55.6	16.8	42.5	10.9	26.7	6.8
8	43.9	9.8	45.8 --	11.8 --	38.5	23.0 --	25.0	8.1
9	49.1	14.2 --	57.5	16.0 --	48.6	23.4 --	32.8	11.2
10	88.1	41.9	108.3	49.5	83.1	37.9	41.5	11.1
11	120.3	48.0	149.1	52.0	111.2	37.8	48.7	19.6
12	150.0	51.3	206.5	76.1	127.6	53.1 ++	49.5	32.6 ++

TABLE 3.2

SUMMARY OF SYNTHETIC MEANS AND ST DEVS FOR 40 YEARS DATA,

DATA SET 2

	SITE NUMBER							
	1		2		3		4	
MONTH	MEAN	ST DEV	MEAN	ST DEV	MEAN	ST DEV	MEAN	ST DEV
1	145.5	50.6	260.4	98.8	133.1	68.4 ++	39.0	36.8 ++
2	124.8	39.5	265.7	90.2	103.7	38.8	25.6	13.6 --
3	98.6	24.4	206.2	64.7	83.2	28.0	23.0	6.2 --
4	105.3	58.3	190.2	126.3 ++	92.7	47.0	29.4	14.9 ++
5	77.9	23.8 --	105.6	43.8	75.1	36.2	37.0	14.6
6	52.8	13.8	60.0	16.8	53.0	16.5	30.6	7.4
7	44.1	11.6	49.7	16.9	41.3	12.0	26.4	6.7 --
8	46.5	10.8	52.7	16.0	49.6	29.8 --	28.5	7.8
9	54.6	20.9	65.2	21.5	63.0	42.5 ++	37.0	14.6
10	86.6	30.2 --	104.0	43.3	82.7	26.3 --	42.1	7.7 --
11	119.3	34.4	153.5	48.9	117.0	42.7	48.2	16.7
12	154.1 ++	42.8	227.4	71.2	141.9 ++	53.2 ++	54.8	31.4 ++

TABLE 3.3

SUMMARY OF SYNTHETIC MEANS AND ST DEVS FOR 40 YEARS DATA,

DATA SET 3

	SITE NUMBER							
	1		2		3		4	
MONTH	MEAN	ST DEV	MEAN	ST DEV	MEAN	ST DEV	MEAN	ST DEV
1	139.3	33.9 --	249.7	61.7 --	118.4	36.5 --	35.1	16.3 --
2	128.2	48.8	271.5	105.6	112.0	49.0	33.3	24.6
3	104.5	28.8	206.4	55.3	91.0	31.6	24.3	8.6 --
4	98.2	47.6	166.3	80.7	72.3	33.0	24.7	9.6
5	74.4	33.0	92.8	46.0	69.5	34.6	36.2	19.0
6	52.7	15.9	60.2	16.0	54.4	19.7	30.6	9.0
7	47.1	11.6	52.5	14.6	41.5	9.9 --	26.2	7.1
8	45.1	11.2	49.2	14.4	39.9	26.4 --	26.1	8.3
9	54.0	20.6	62.0	23.2	57.1	33.6	36.3	14.1
10	90.7	41.3	102.2	41.7	85.5	41.2	42.0	14.1
11	114.8	36.1	142.6	53.2	113.4	38.9	46.7	15.4
12	132.0	31.7 --	190.0 --	50.1 --	121.0	34.9	46.5	28.4

TABLE 3.4

SUMMARY OF SYNTHETIC MEANS AND ST DEVS FOR 40 YEARS DATA,

DATA SET 4

	SITE NUMBER							
	1		2		3		4	
MONTH	MEAN	ST DEV	MEAN	ST DEV	MEAN	ST DEV	MEAN	ST DEV
1	140.4	44.2	248.2	70.8	120.1	42.4	33.8	17.2
2	131.4	50.3	277.0	123.8	105.4	50.5	30.5	25.2
3	104.7	29.3	207.5	63.6	85.4	33.7	23.9	8.1 --
4	89.0	43.3	161.5	96.6	77.9	41.5	22.5	11.4
5	72.8	25.6	105.5	49.6	76.1	34.0	33.9	14.5
6	52.8	14.9	61.6	18.2	54.7	17.5	30.9	7.5
7	48.0	13.8	56.9	18.2	43.9	13.4	27.5	8.5
8	45.0	9.2 --	51.8	16.0	45.8	28.3 --	26.4	9.5
9	48.7	15.1	62.5	22.3	50.7	28.7	31.5	9.5 --
10	92.5	40.2	111.7	53.6	86.2	35.2	42.2	11.3
11	117.1	40.2	150.1	56.0	114.3	39.9	44.5	15.0
12	141.4	46.0	200.7	64.4	122.5	37.8	44.0	22.7

TABLE 3.5

SUMMARY OF SYNTHETIC MEANS AND ST DEVS FOR 40 YEARS DATA,

DATA SET 5

	SITE NUMBER							
	1		2		3		4	
MONTH	MEAN	ST DEV	MEAN	ST DEV	MEAN	ST DEV	MEAN	ST DEV
1	151.4	46.5	272.9	80.2	136.2	49.3	42.0 ++	24.4
2	148.3 ++	50.6	306.9 ++	102.3	123.9 ++	45.0	32.6	15.4 --
3	99.7	26.3	193.7	48.0	85.2	35.2	26.1	12.4
4	94.2	43.3	149.1	68.8 --	80.8	37.0	26.8	12.2
5	77.3	29.4	95.6	37.2	71.7	37.7	36.1	21.5 ++
6	53.7	13.0	63.2	16.6	54.8	18.6	30.9	9.2
7	46.8	13.8	52.3	14.3	43.5	12.9	27.2	9.3
8	44.9	13.3	47.0	17.4	44.0	36.8	25.9	8.7
9	42.2 --	14.7	50.8 --	16.5 --	39.3 --	21.0 --	28.4 --	9.3 --
10	83.3	39.1	99.7	51.9	76.3	37.4	39.2	12.1
11	112.0	30.7 --	151.3	45.1	112.5	34.9	46.4	17.2
12	147.4	38.7	228.1 ++	65.5	134.2	35.3	49.4	21.2

TABLE 3.6

SKEWNESS AND LAG-ONE SERIAL CORRELATION COEFFICIENTS FOR HISTORIC AND SYNTHETIC SERIES

Station Number	Skewness Coef. C_S						Lag-one Serial Correl. Coef. r(1)					
	Historic Series	Synthetic Series					Historic Series	Synthetic Series				
		1	2	3	4	5		1	2	3	4	5
1	0.88	0.99	0.58	0.81	0.88	0.72	0.34	0.40	0.32	0.34	0.37	0.37
2	0.71	1.11	0.87	0.84	1.07	0.74	.26	.31	.29	.27	.28	.28
3	1.64	0.87	0.93	0.78	1.05	1.15	.38	.34	.34	.46	.34	.40
4	0.97	1.00	0.97	1.03	1.12	1.04	.53	.45	.34	.45	.31	.48

TABLE 3.7

LAG-ZERO CROSS CORRELATION MATRICES FOR HISTORIC AND SYNTHETIC SERIES

Historic	1.00			
	.87	1.00		
	.75	.79	1.00	
	.71	.67	.80	1.00
Synthetic Series 1	1.00			
	.84	1.00		
	.79	.80	1.00	
	.72	.65	.80	1.00
Synthetic Series 2	1.00			
	.84	1.00		
	.76	.75	1.00	
	.69	.61	.76	1.00
Synthetic Series 3	1.00			
	.84	1.00		
	.78	.78	1.00	
	.68	.60	.78	1.00
Synthetic Series 4	1.00			
	.85	1.00		
	.76	.80	1.00	
	.68	.64	.77	1.00
Synthetic Series 5	1.00			
	.84	1.00		
	.75	.79	1.00	
	.66	.66	.78	1.00

TABLE 3.8

ANNUAL FLOW STATISTICS FOR HISTORIC AND SYNTHETIC SERIES

Station Number	Mean Annual Flows CSD x 1000						Annual Std Devs CSD x 1000					
	Historic Series	Synthetic Series					Historic Series	Synthetic Series				
		1	2	3	4	5		1	2	3	4	5
1	1078	1049	1110	1081	1083	1101	173	156	139	142	164	174
2	1652	1570	1630	1622	1649	1688	270	234	297	228	243	237
3	981	938	1034	985	979	1001	195	156	173	176	163	171
4	387	400	419	412	393	392	73	89	77	83	66	68

Station Number	Skewness coefs. of annual flows						Lag-one Serial correl. coefs. of ann.flows					
	Historic Series	Synthetic Series					Historic Series	Synthetic series				
		1	2	3	4	5		1	2	3	4	5
1	.27	-.19	-.01	.51	-.10	.43	.06	.26	-.03	0.26	.03	.12
2	.45	.06	.50	.25	.54	.09	-.07	.25	.00	.05	-.29	-.04
3	.99	-.10	-.38	.36	.70	.42	.04	.06	-.11	-.14	.01	.23
4	1.43	-.08	.73	.52	-.01	.29	-.05	.13	-.14	0.32	-.21	.13

the natural data are normally distributed. The result for the 5-40 years series is that 94% of the standard deviations lie in the range of $\pm$ two standard errors, but skewness values are not even approximately preserved.

A reason for the log-normal distribution apparently failing to reproduce parameters within given confidence limits is that estimates of means and standard deviations for log-normal distributions are only approximately normally distributed (Brooks and Carruthers, 1953). The confidence limits established require the estimates to follow a normal distribution. This would account for at least some of the reduction to only 75% of the estimates lying within 95% confidence limits.

An alternative form of confidence test which does give satisfactory results is described in section 2.6 of Chapter 2.

3.5 SUMMARY

Several multi-stream flow generating methods were described. A multivariate technique due to Matalas (1967b) has been implemented and results from applying the scheme to four flow stations in the Waitaki river basin are given.

Good resemblance between historic and synthetic means, correlations and skewnesses has been achieved. Resemblance between historic and synthesized standard deviations was not as close as expected.

With this limitation, the model is used to generate input data for a month-by-month operation study of a model of a hydro-electric power system, part constructed and part proposed, for the Waitaki river system.

CHAPTER FOUR

OPERATION OF A SINGLE-STRUCTURE SINGLE-PURPOSE SYSTEM

4.1 INTRODUCTION

The concept of implicit stochastic optimization for storage operation by using deterministic dynamic programming coupled with streamflow simulation has been described in Chapter 1. By describing the derivation of rules for operating a single irrigation storage, this chapter gives an introductory example.

The Manorburn Dam in Central Otago was selected as a real example for this study, because a good hydrologic record was available and because there was a "real-time" operating problem. This "real-time" problem involved making a decision at the beginning of each irrigation season on the volume of water to be released during the season, and how much to store for future seasons.

Compared with most water resource systems, this single-purpose single-structure system is very simple, and other optimizing techniques, particularly linear programming and certain forms of non-linear programming, could equally well be used to derive operating policies. However, this system provides a good introduction for demonstrating the application of deterministic dynamic programming in conjunction with streamflow simulation.

4.2 DESCRIPTION OF SYSTEM

4.2.1 Storage Capacity

The capacity of the Manorburn Dam is 41,700 acre feet (ac.ft). For this study the active storage volume is taken as 40,000 ac. ft. This represents a very conservative design, for the dam overflows very infrequently. Spillage was only once recorded in the period 1942 to 1956.

4.2.2 Inflows, Evaporation and Seepage

Records of net annual inflows to the dam were derived from records of reservoir levels and releases for 48 years, from 1918 to 1965. Evaporation and seepage losses were not included to give total inflows, it being assumed that the losses which occur under any derived operating policy would be similar to those which have actually occurred during the period of record.

For these 48 years, the net inflow parameters were:

Mean inflow	$\bar{Q}$	=	19,000 ac.ft
Std. deviation	S	=	7,740 ac.ft.
Skewness coeff.	C_s	=	0.85
Lag one serial correl. coeff.	$r(1)$	=	0.08

The close fit of the data to a log-normal distribution is illustrated by Figures 4.1 and 4.2. The serial correlation is not significant and the flow series can be represented as a series of log-normally distributed random variates with appropriate mean and variance.

4.2.3 Area Irrigated

The dam stores water for the 12,360 acre Ida Valley irrigation scheme. In addition, some water is diverted to the neighbouring 2,660 acre Galloway scheme. This amount is conditional on the volume in storage in the smaller Poolburn Dam. Compared with the total annual release, this diversion is small and it is assumed that the Manorburn is operated independently of the Poolburn. To compensate, the area irrigated will be taken as 14,000 acres.

4.3 ALLOCATION OF WATER

4.3.1 Quotas

Water is allocated to irrigators by quotas. The quota is the amount of water the reservoir operator contracts to supply to an irrigator each year. In normal years, irrigators pay for these quotas. In this exercise, total quotas are taken as 14,000 ac.ft of water per year; this is equivalent to 12 inches

FIG. 4.1 ANNUAL INFLOWS TO MANORBURN DAM, 1918-1965. NORMAL PROBABILITY PLOT

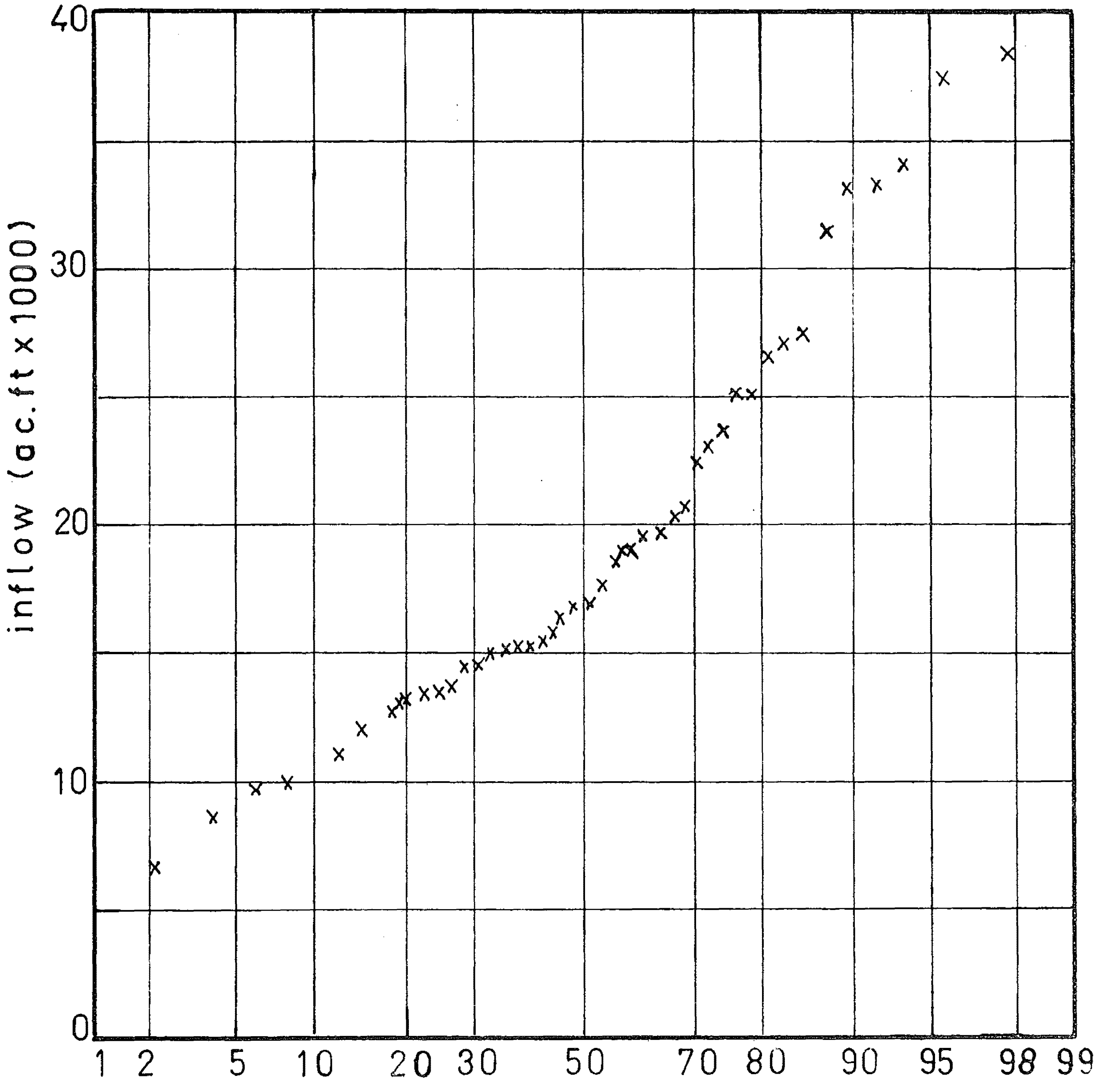
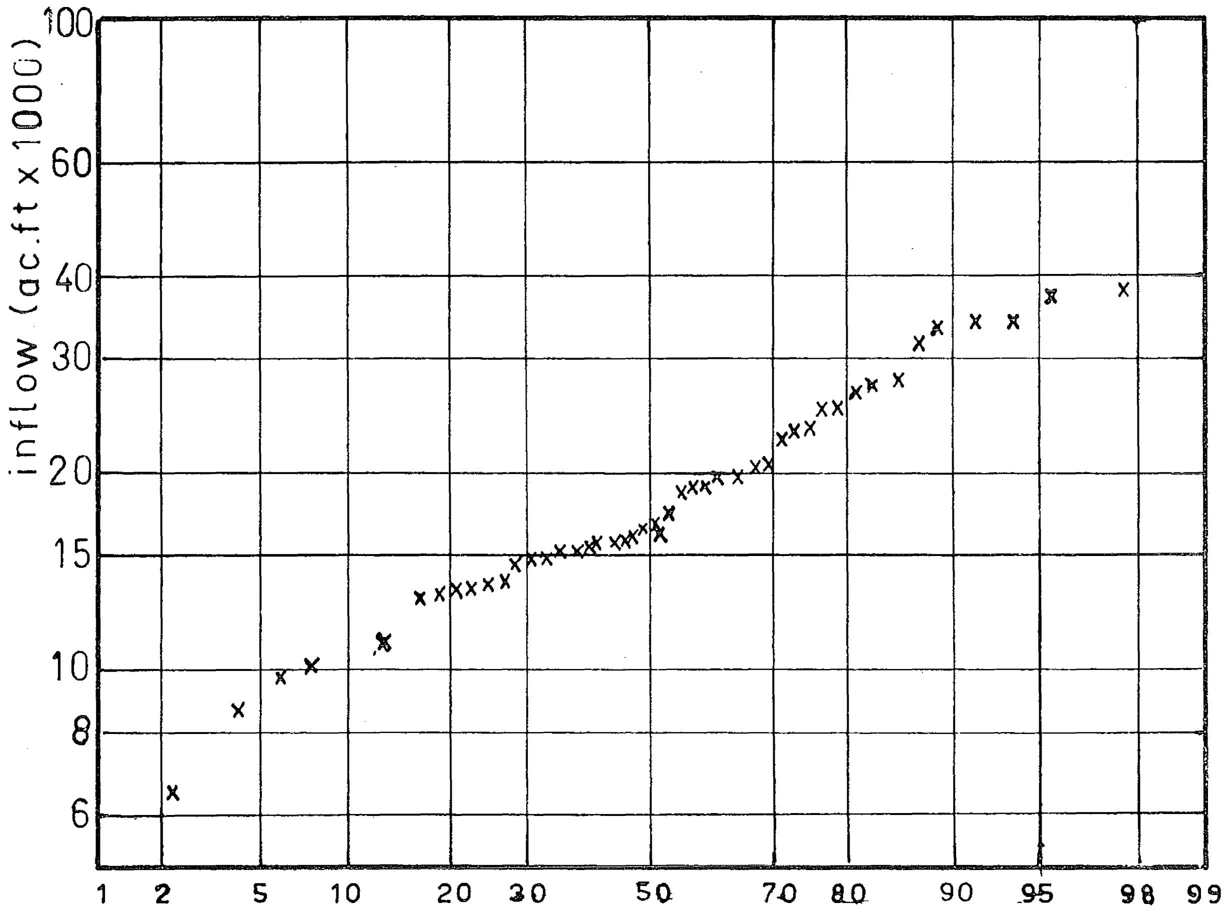


FIG. 4.2 ANNUAL INFLOWS TO MANORBURN DAM, 1918-1965. LOG-NORMAL PROBABILITY PLOT



of water per year over all the irrigated area. 14,000 ac.ft. is the volume supplied to irrigators. Because of losses in the distribution network, the volume which must be released at the dam to supply full quotas is much greater. If these losses are assumed to be a constant 25% of the release at the dam, the release required at the dam to supply full quotas is 18,700 ac.ft.

4.3.2 Rationing

In dry years, the dam may be down some at the beginning of the irrigation season and it may be prudent for the operators to ration supplies of water to a percentage of the full quotas. In such a year irrigators in this scheme pay only for that part of the quota supplied.

4.3.3 Sales

In wet years, water in excess of quotas may be available, and in this scheme it may be sold for half the quota rates. Such water is known as sales. In any one year, total sales will be constrained to a maximum of 30% of the quotas, since plentiful years at the storage when sales are likely, are also likely to be good years for irrigators, who will not be seeking large additional supplies. Thus the maximum useful release from the dam will be 130% of the release required to supply full quotas, that is, 24,300 ac.ft.

4.3.4 Return Function

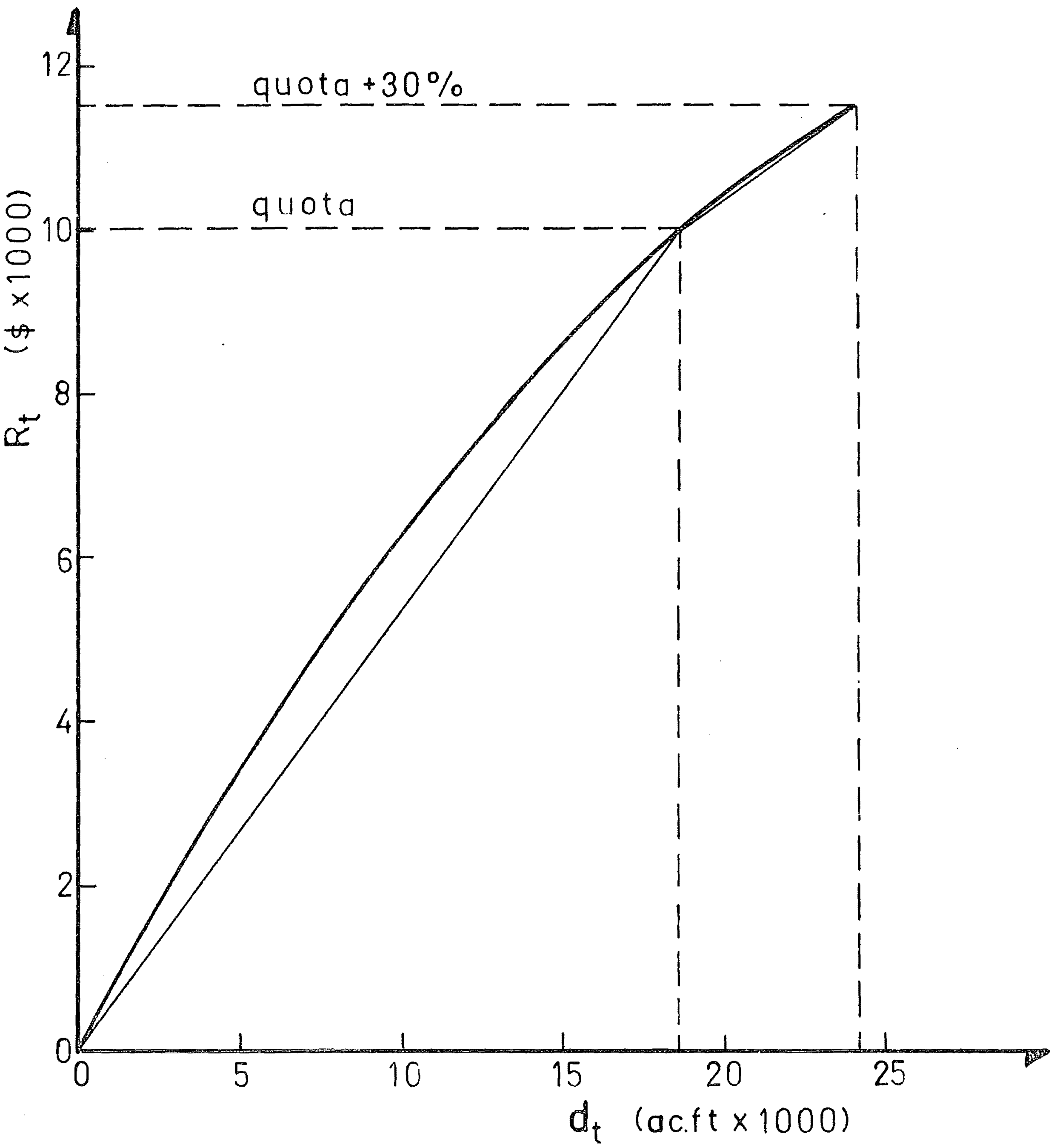
On the basis of the information given above, a bilinear function may be prepared to show the annual return received by the reservoir operator for any release from the storage. This function is shown by the two straight lines in Figure 4.3.

This can be described as a short term benefit loss function, as distinct from a long term benefit function which at the planning stage would represent income from different levels of quotas allocated for the scheme.

The short term return function is of great importance in operation studies. It means that a monetary value can be imputed to every possible level of release made in a particular time interval. A function of this type

FIG. 4.3 MANORBURN DAM, PROPOSED QUADRATIC BENEFIT FUNCTION

$$R_t = 0.739\, d_t - 1.09 . 10^{-5}\, d_t^2$$



is required in any operation optimizing study.

In the present example, a straight forward definition is possible because of the arrangement for charging for supplies below and above the quota levels. In many other examples, and in other uses of water, such as hydro-electric power generation, an explicit definition of this type is not always possible.

For the purpose of optimizing operation, an indirect definition can sometimes be achieved. Chapter 5 will give an example of this.

4.4 OBJECTIVES OF OPERATION

With a return function thus defined, the objective for the reservoir operators over a period of time will be to maximize the present value of returns less operating costs.

In fact, for all the Central Otago irrigation schemes, revenues fail to cover operating costs, and the schemes are supported by substantial government subsidies. In this exercise, operating costs will be taken as constant and independent of the volume of water released; thus the objective will be to maximize the present value of revenues. Revenues are understood to be the returns given by the function in Figure 4.3. In this way, all the benefit from operating the storage in a particular way over a number of years can be expressed as a single valued scalar quantity, namely the present value of returns.

The bilinear function in Figure 4.3 has a weakness, in that it cannot always differentiate between releasing water and storing it for release later. This leads to some irrational results.

For instance, in dry years it is intuitively more desirable for irrigators to receive say 85% quotas for two successive years rather than 75% in one year and 95% in the next, but in terms of benefits, the bilinear function ranks both of these situations equally. Similarly in plentiful years, two successive years of quotas plus 10% sales should be ranked ahead of quotas only in one year and quotas plus 20% sales in the other.

This difficulty can be overcome by postulating a continuous curve in place of the bilinear function. Many studies have used quadratic functions of the type;

$$R_t = a\, d_t^{\,2} + b\, d_t + c \qquad (4\text{-}1)$$

where R_t is the return for the tth year.

d_t is the release made during the tth year

a, b and c are constants.

This function, fitted to the Manorburn data, is shown as the smooth curve in Figure 4.3. This gives a close approximation to the bilinear function and has the desirable property of decreasing increments in returns for extra units of water sold. This can also be described as decreasing marginal returns.

4.4.1 Operating Rules - the Present Policy

In operating the Manorburn Dam, decisions must be made at the beginning of each irrigation season on the amount of water to release to irrigators during the season. Irrigators are notified of the proportion of their quotas and possible sales that will be available during the ensuing season and can plan their operations accordingly.

This decision is important, for it is the key to successful operation of the dam. If too much water is released, the storage may not fill sufficiently for the next season and short supplies and accompanying reductions in returns will result. If too little is released, large amounts of water may be sold as sales in the following years and possibly some will be spilled if the dam fills. Clearly, there is a need for a defined operating rule which indicates an optimum release.

In practice, releases to be made from the Manorburn Dam during irrigation seasons are decided in the following way.

At the beginning of each season, in September, the next two years of operation are considered. The volume in storage at the start of the season and the expected inflow during the next 12 months are summed, and this is considered to be the total quantity of water available for use during the next two seasons. This water is divided equally between the two years. Thus the anticipated release during the irrigation season is -

$$d_t = \tfrac{1}{2}(s_t + \bar{Q}) \tag{4-2}$$

where d_t is the anticipated release during the tth year,

s_t is the volume in storage at the start of the tth year,

$\bar{Q}$ is the expected annual inflow, which estimated from 48 years data, was 19,000 ac. ft.

This rule, (4-2), will be referred to as the "present policy".

4.5 DERIVATION OF OPERATING RULES

4.5.1 Dynamic Programming Approach

For the Manorburn Dam, only one decision per year was required and an analysis on a year-by-year basis is permissible. An irrigation year is from 1st September to 31st August. An irrigation season starts in September when irrigators are notified of the proportion of quotas and possible sales which it is anticipated will be available for the impending irrigation season, from September to April.

It was assumed that only half the inflow during an irrigation year would occur in the first eight months of the year when it could be used in the irrigation season of that year. The remaining half occurred in the following four winter months and was stored for the following season. Thus the maximum release in a particular irrigation season was constrained to be less than the initial storage plus half the inflow.

$$\text{i.e.}\quad d_t \leqslant s_t + \tfrac{1}{2}\bar{Q} \tag{4-3}$$

Writing equation (1-7) in the notation of this chapter, and dropping the term for losses -

$$f_n(s_t) = \max\left(R(d_t) + \frac{1}{1+r} f_{n-1}(s_t + q_t - d_t)\right), \qquad (4\text{-}4)$$

subject to, $0 \leqslant d_t \leqslant \min\left(D_M, \left(s_t + \frac{\bar{Q}}{2}\right)\right),$

and, $0 \leqslant s_t \leqslant S_M.$

To derive a real-time operating policy, the steps summarized in Figure 1.3 are followed. If (4-4) is solved for a particular set of inflow data, the result is a policy which maximizes the present value of annual returns, should a particular set of inflows occur.

A mean policy for real-time operation is obtained by finding a best fit function of the type

$$d_t = g(s_t, q_{t-1}) \qquad (4\text{-}5)$$

The approach assumes that the "steady-state" operation of the system is under study; that is the objectives are unchanged from one year to another and the system is operated over an infinte time period. If the analysis is made on a monthly basis, it is necessary to derive a set of functions of the type (4-5).

Persistence between annual flows into the Manorburn is negligible and q_{t-1} does not contribute any information towards the magnitude of q_t. Apart from the possibility of predicting q_t by snowpack data, s_t is the only variable describing the condition of the system at the beginning of year t. Snowmelt prediction is not investigated here, and the problem is reduced to finding the best fit function.

$$d_t = g(s_t) \qquad (4\text{-}6)$$

4.5.2 Shortage Index

Because of the constraint (4-3), occasions occur in real-time operation with rules (4-2) and (4-6) when the storage level is low at the start of the season, and expected inflows during the season do not eventuate. In some of these situations, anticipated releases to irrigators, predicted at the beginning of the season under the quota system, are not available. In such cases, releases were set equal to what was available, leaving the storage empty at the end of the irrigation season.

Apart from evaluating returns on the basis of actual deliveries, this failure to deliver anticipated supplies is not penalised by the return function. For this reason a quantity I_s, the shortage index, is defined as -

$$I_s = \sum_{t=1}^{N} \left[\frac{d_t^{antic} - d_t^{act}}{d_t^{antic}} \right]^2 \cdot \frac{100}{N} \qquad (4\text{-}7)$$

$$\text{for all } d_t^{act} \leqslant \text{Quota}$$

$$1 \leqslant t \leqslant N$$

where N is the length in years of the period analysed.

d_t^{antic} is the anticipated release at the start of the tth irrigation year

d_t^{act} is the actual release.

Effectively, this is the sum of squares of the unavailable fractions of anticipated releases, adjusted for a base length of 100 years, for all releases less than 100% quota.

This index ranks shortages less than quota which were not anticipated according to the square of their magnitude relative to the respective anticipated release. Were they available, it is expected that losses to irrigators due to unanticipated shortages would be approximately proportional to the shortage index.

The index is not involved within the optimizing process, rather it is included here to give a fuller description of the results obtained by using different operating policies.

4.5.3 Model Specification

The preceding sections of this chapter have described a simple irrigation storage, a method for deriving a policy for its year-by-year operation, and measures of the effectiveness of this policy.

It remains to describe the numerical procedures followed and the results obtained.

4.6 MODEL SOLUTION

4.6.1 Testing Procedure

(1) Based on the statistics of the 48 years historic data, 5-48 year and 1-600 year sets of inflow data were generated by sampling from a log-normal distribution. Parameters describing these sets are given in Columns 2 to 8 of Table 4.1.

(2) Assuming that annual returns are discounted to their present value at a rate of 5% per annum, direct dynamic programming solutions were carried out for these data sets,as well as the historic data, to determine the best possible operating policies and the maximum average annual returns, consistent with the objectives and constraints outlined previously.

To ensure that these were steady-state solutions, unaffected by terminal state optimization, an additional 24 years of synthetic data were added to the end of each of the data sets.

(3) Excepting results from these last 24 years, which were discarded, shortage indices and annual returns were evaluated.

(4) Multiple linear regression was used to fit functions of the type (4-6) to these seven sets of results.

(5) With the policy obtained from the dynamic programming analysis of the historic data, storage behaviour analyses, simulating real-time operation, were made for the 5-48 year sets of synthetic data.

(6) For comparison, the present policy was used in storage behaviour analyses of the same five synthetic data sets.

(7) Sensitivity analyses:

(a) Direct dynamic programming solutions were obtained for the historic data assuming r = 0%, 10%.

TABLE 4.1

MANORBURN DAM

STORAGE BEHAVIOUR ANALYSES - FLOW PARAMETERS AND ANNUAL RETURNS

Flow Data								Annual Returns (Dollars)					
Flow Description			Flow Parameters					Av. Annual Returns			St.devs annual returns		
									Real-time operation			Real-time operation	
File Name	Length L (yrs)	Type	Mean ac.ft.	st.dev. ac.ft.	coef. var.	coef. skew.	Lag one serial correl.	Determin. soln.	Optimum linear policy	Present policy	Determin. soln.	Optimum linear policy	Present policy
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
MANBH	48	Historic	19000	7700	0.41	0.85	0.08	10020	-	-	930	-	-
MANB6	600	Syn.	19200	7600	0.40	0.00	-0.01	10100	-	-	900	-	-
MANB1	48	Syn.	17600	7500	0.42	0.90	0.11	9540	9470	9380	960	1320	1420
MANB2	48	Syn.	19800	7600	0.39	0.59	0.10	10140	10060	10010	1060	1050	1310
MANB3	48	Syn.	19700	7100	0.36	0.65	0.08	10200	10020	10170	1320	1040	1280
MANB4	48	Syn.	18500	6900	0.38	0.85	-0.12	9870	9860	9770	960	1100	1220
MANB5	48	Syn.	18600	6600	0.35	0.65	-0.28	9940	9950	9860	740	770	960

(b) To ascertain the effect of serial correlation in the inflows, 600 year long data series with the same means but different degrees of serial correlation were generated and used in storage behaviour analyses with the derived policy.

(c) To investigate the result of incorrectly estimating the mean annual flow, data series were generated and storage behaviour analyses carried out assuming that the estimated means were in error by -10%, -5%, 0%, 5% and 10%.

(d) Analyses similar to (c) were made assuming errors of estimate in the standard deviation of annual flows of -30%, -15%, 0%, 15% and 30%.

The computer programs developed for this work are described in Appendix I.

4.6.2 Presentation of Optimal Policy

Regression results giving a functional form of optimal policy are in columns 2 to 10 of Table 4.2. To illustrate the derivation of the policy function, optimum releases obtained from a dynamic programming analysis of the historic data are plotted with initial storage volumes in Figure 4.4, and the best fit linear function is shown. In Figure 4.5 this function is redrawn, together with the present policy. In this situation with a quadratic return function, linear policy functions provided as good a fit to the optimal data as did more complex logarithmic and quadratic functional forms (ref. columns 2 to 10 of Table 4.2). This result was also observed by Young (1966).

The linear policy functions obtained for the historic data and the 600 years synthetic data were the same, while the policies obtained for the 5-48 year synthetic data series were very close. If a real-time operating policy, derived by using the least squares criterion to fit a functional form to the results of a direct dynamic programming solution, can be described as optimal, the linear policy derived for either the historic data or the 600 years of synthetic data may be described as the "optimal linear policy".

TABLE 4.2

MANORBURN DAM

STORAGE BEHAVIOUR ANALYSES - POLICY FUNCTIONS AND SHORTAGE INDICES

Flow Data	Policy Functions									Shortage Index I_s		
File Name	Linear function $d_t = a_o + a_1 s_t$				Quadratic function $d_t = a_o + a_1 s_t + a_2 s_t^2$					Deterministic solution	Real-time operation	
	a_o	a_1	Multiple Correl. R	Std. error	a_o	a_1	a_2	R	S.E.		Optimum linear policy	Present policy
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
MANBH	16800	0.16	0.38	2700	13800	0.575	-0.0011	0.46	2600	0.013	-	-
MANB6	16800	0.16	0.42	2500	15900	0.283	-0.0003	0.43	2500	0.017	-	-
MANB1	14900	0.18	0.50	2200	-	-	-	-	-	0.016	2.617	0.035
MANB2	17000	0.16	0.36	2900	-	-	-	-	-	0.006	1.135	0.064
MANB3	17400	0.15	0.33	2700	-	-	-	-	-	0.010	1.011	0.013
MANB4	16900	0.20	0.40	2600	-	-	-	-	-	0.010	1.416	0.030
MANB5	16000	0.18	0.33	2100	-	-	-	-	-	0.021	0.681	0.000

FIG. 4.4 DERIVATION OF LINEAR REAL-TIME POLICY FROM DETERMINISTIC SOLUTION FOR 48 YEARS HISTORIC DATA, r = 5%

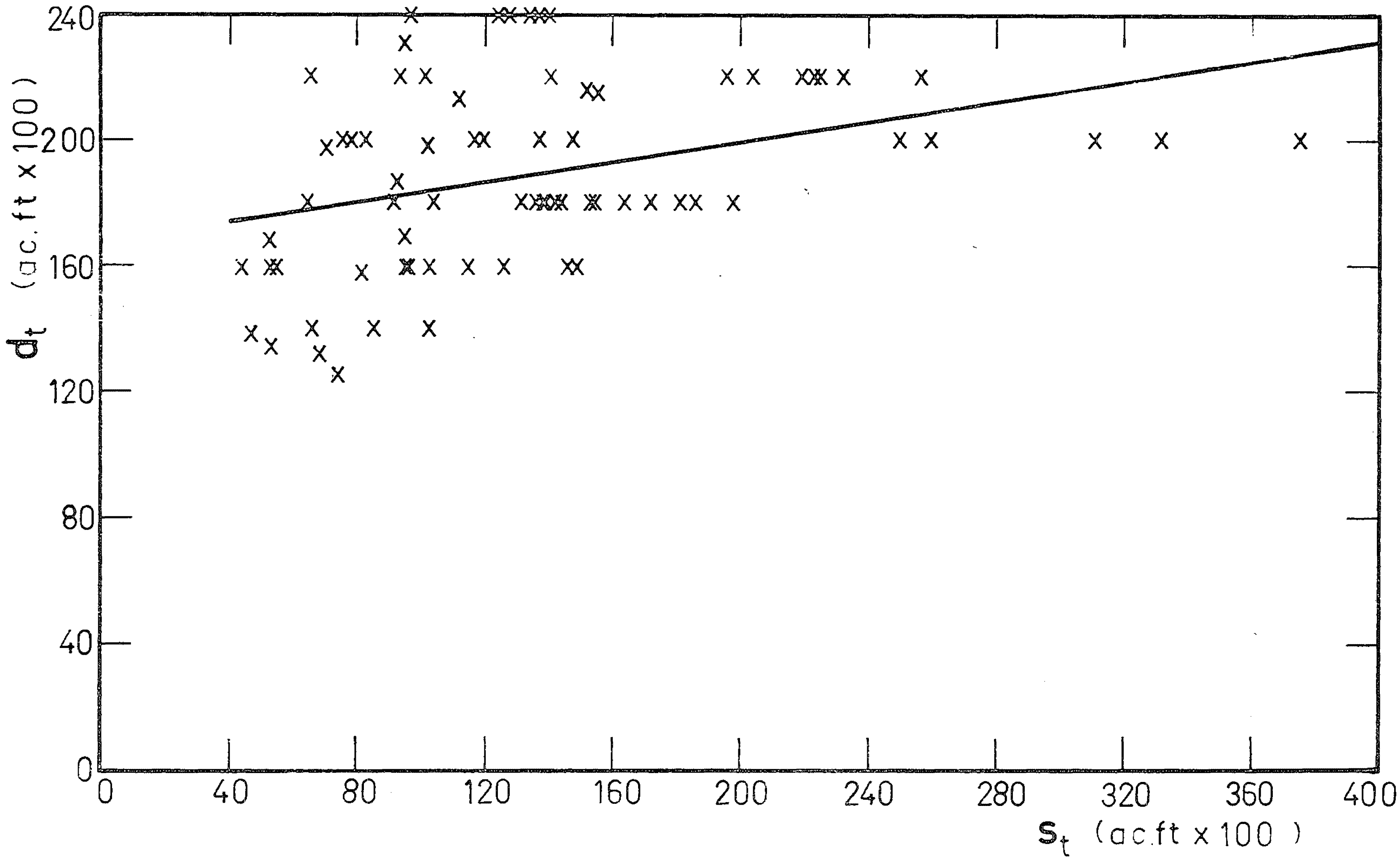
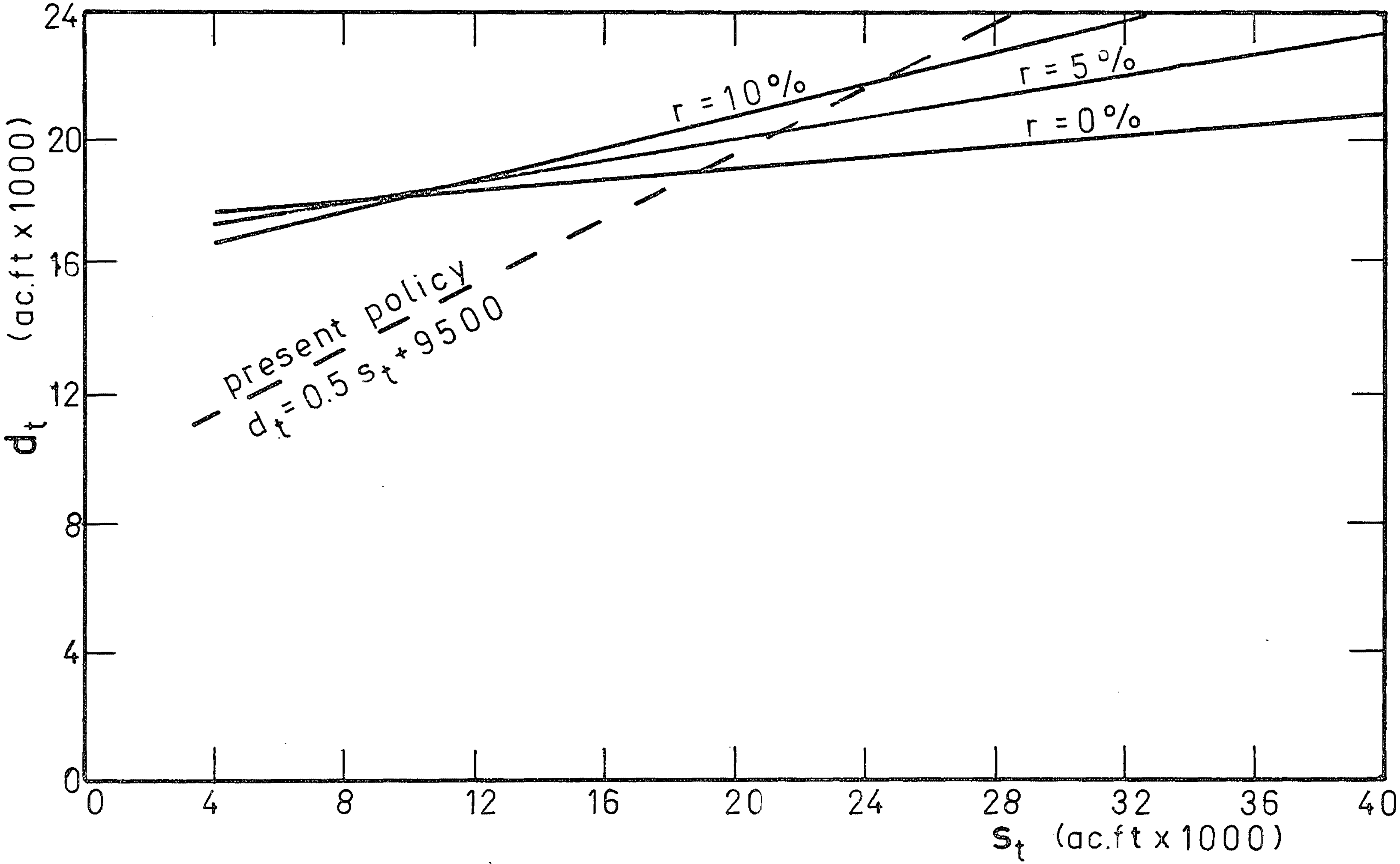


FIG. 4.5 COMPARISON OF OPTIMAL LINEAR POLICIES FOR r = 0, 5, & 10 % & ALSO THE PRESENT POLICY



This real-time policy is the best available estimate of the true and unknown population linear policy. Because this policy has been derived directly from the historic data, it would appear that in this situation, the synthetic hydrology concept is redundant.

4.7 COMPARISON OF OPERATING POLICIES

4.7.1 Results of Storage Behaviour Analyses

Although maximization of the present value of annual returns was used as the criterion to derive the linear policy, average annual returns are of greater use in assessing the relative merits of different policies.

Values of average annual returns are estimates of the population value implied by the use of the policy derived from the parent population for the particular functional form (in this case linear) for which least squares fit is obtained. Thus two types of error are present in estimates of the average annual returns;

(1) error from using an estimate of a policy, rather than the population policy.

(2) error involved in the calculation of returns from a particular flow sequence given a fixed policy.

Because of the close similarity of policy estimates derived from different data series, it is concluded that this type-one error must be very small.

The type-two error is best examined by considering the results of the analysis with the optimum linear policy. The results, in terms of average annual returns and standard deviations of annual returns are given in columns 10 and 13 of Table 4.1. These average annual returns appear to be proportional to the mean annual flows given in column 4 of Table 4.1. The mean of the five values for average annual return is $9870.

Given that a particular set of inflows will occur, the absolute optimum policy for operation is defined by the direct dynamic programming solution.

This deterministic solution requires the inflows to all be known at the start of operation.

To check how closely the derived linear real-time policy approximates the absolute optimum, and how closely the present policy follows the derived real-time policy, returns for operation with the 5-48 year sets of inflows are tabulated in Table 4.1. Average annual returns for (1) the direct dynamic programming solution, (2) the derived linear real-time policy and (3) the present policy are listed in columns 9, 10 and 11 respectively. Standard deviations of these annual returns are listed in columns 12, 13 and 14 respectively, and shortage indices for the three solutions are in columns 11, 12 and 13 of Table 4.2.

These returns from using different policies are all close. On average, the maximums obtainable, given by the dynamic programming solutions, are 0.7% greater than the results using the optimal linear policy. In turn, the optimal linear policy results are on average 0.4% greater than the results for the present policy.

4.7.2 Shortage and Risk

Results for shortage indices show a wider spread. Values of I_s are close to zero for the dynamic programming solutions. They are not exactly zero because midpoint, rather than lower bound values of storage were taken in forming discrete blocks in the analysis, and occasional small deficits in the volume of water available occur. Similar results show for the present policy analyses, but analyses by the optimal linear policy give much higher values. In effect, use of the optimal linear policy results in a greater proportion of anticipated supplies not being delivered.

Thus, although the optimal linear policy may be the policy to use if the objective is to maximize annual returns, it is not necessarily the best for the Manorburn system. For this situation, the present policy effects a nice compromise in sacrificing a small increment in annual returns to greatly improve the reliability of anticipated supplies.

Reference to Figure 4.5, where the optimal linear policy and the present policy are graphed, illustrates this point. When the storage is less than about half full, the optimal linear policy recommends a greater release than the present policy, while the reverse is true for the storage greater than half full. By recommending a lower release when the storage is low, the present policy indicates a greater aversion to failure of anticipated supplies, as is indicated by the shortage index. This represents a reduction in the risk of failure of anticipated supplies.

This empirical shortage index is one of two distinct risk measures indicated here. It indirectly measures the risk of failure of anticipated supplies. It could be made zero by never anticipating a release greater than the volume in storage at the beginning of the irrigation season, but with the result of drastically reducing the average annual returns.

The second risk measure is a measure of the variability of annual returns. This is given by the standard deviation of annual returns in columns 12, 13 and 14 of Table 4.1. For the optimal linear policy, values are about the same as those for the direct dynamic programming solution, while values for the present policy are somewhat greater. This represents greater variability in releases, a consequence of the steeper gradient of the present policy line shown in Figure 4.5.

The storage behaviour analyses for the first 48 year synthetic data series are illustrated in Figures 4.6 and 4.7. Volumes of water in storage, and inflows are plotted as functions of time. These are on an annual basis; within year storage fluctuations are not shown. The storage never shows empty here because of the constraint (4-3) mentioned earlier. The storage behaviours under the three modes of operation are similar, the curve representing the optimal linear policy tending to be somewhat closer to the dynamic programming optimum than that resulting from use of the present policy.

FIG. 4.6 STORAGE BEHAVIOUR ANALYSES FOR SYNTHETIC DATA FILE MANB1, YEARS 1-24

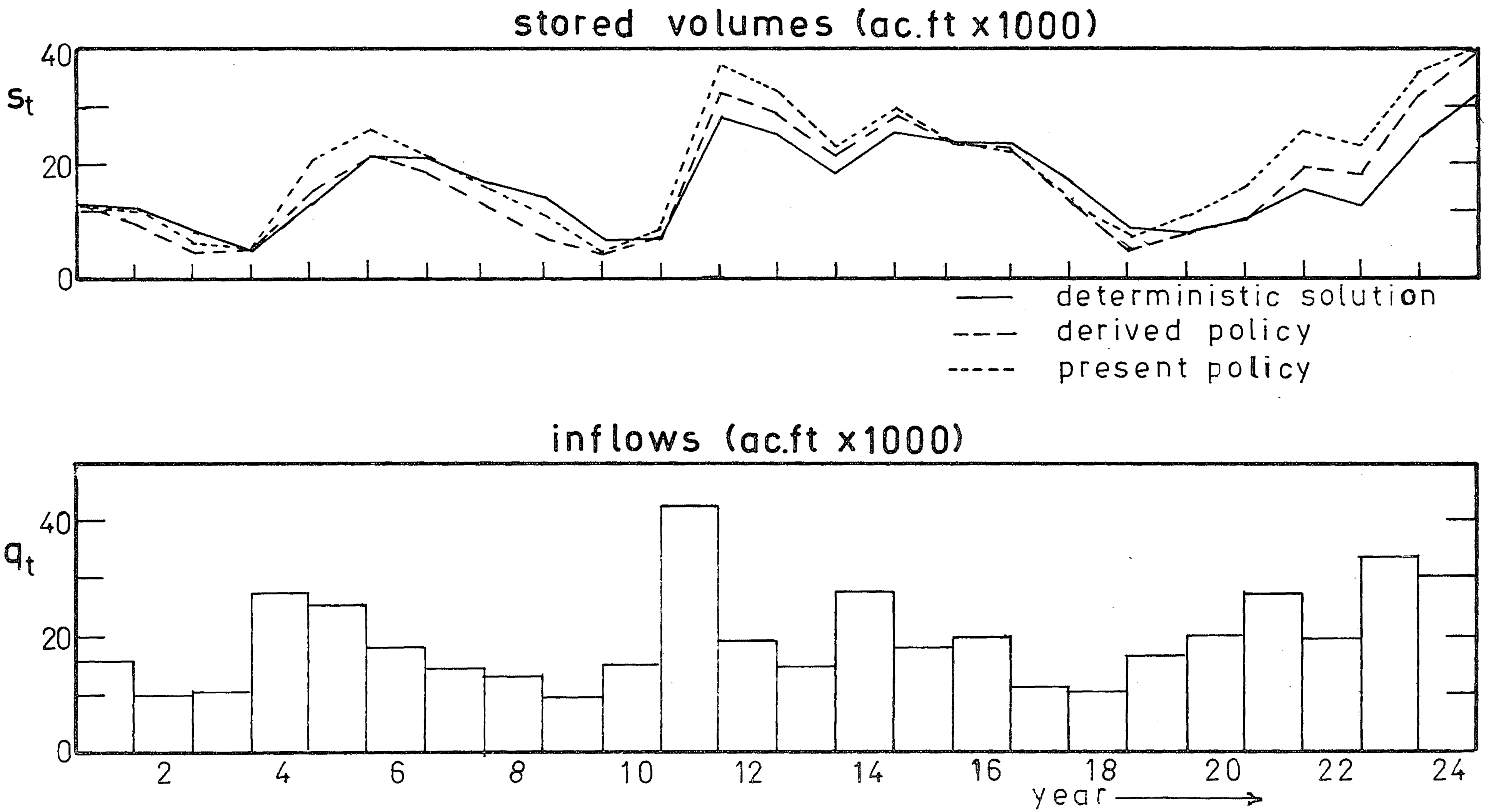
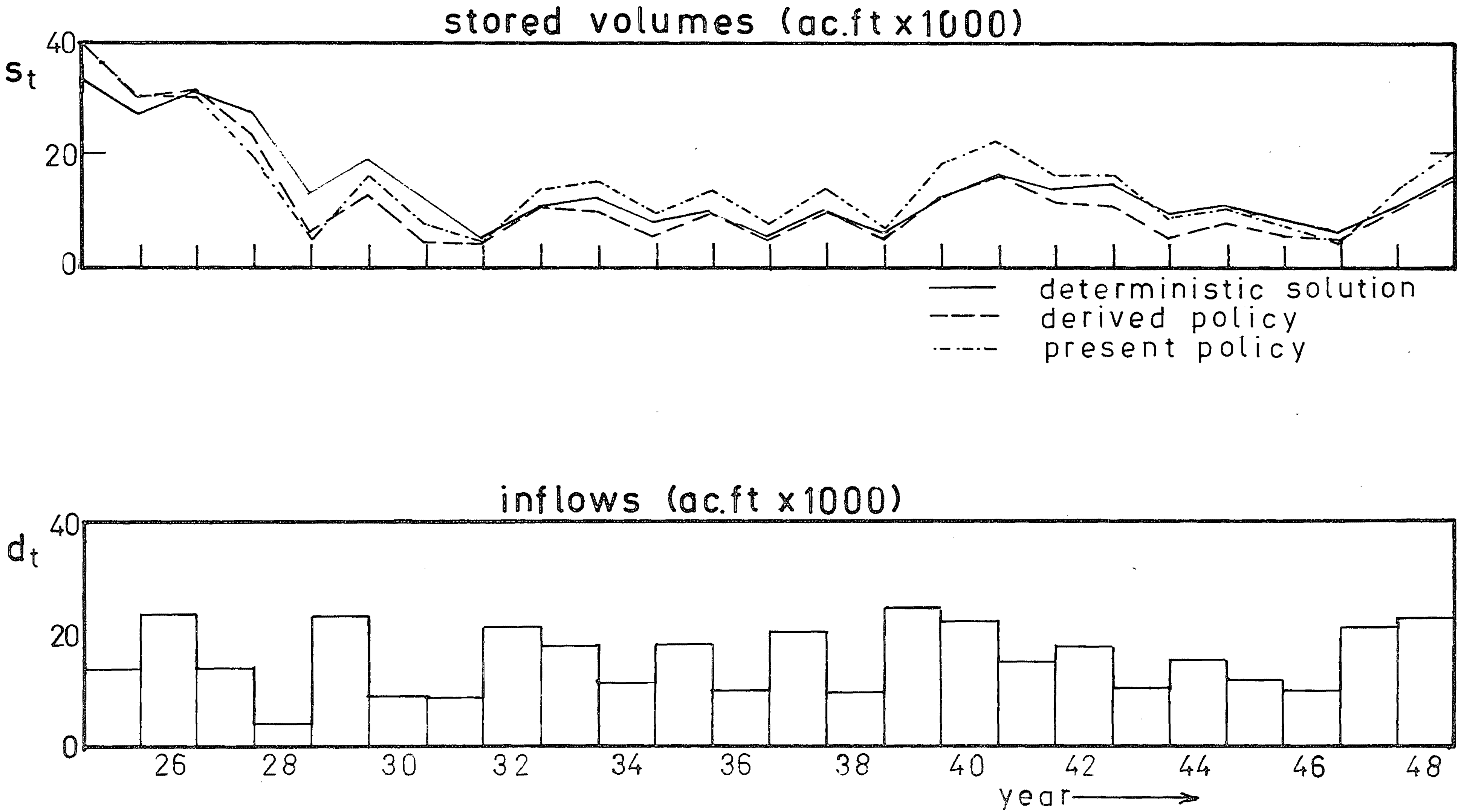


FIG. 4.7 STORAGE BEHAVIOUR ANALYSES FOR SYNTHETIC DATA
FILE MANB1, YEARS 25-48



4.8 SENSITIVITY ANALYSIS

4.8.1 General

To investigate the effect on the operation of the system of incorrect estimates of parameters a simple sensitivity analysis is undertaken. Studied here are results of storage behaviour analyses for different values of the interest rate, the serial correlation between flows, and different estimates of the mean and standard deviation of the flows.

All the results so far described have assumed an interest rate of 5%, and have taken inflows as having a mean = 19000 ac.ft. and standard deviation = 7700 ac.ft. They further assume that annual flows are independent variables following a log-normal distribution. Because a 48 year record is available, these hydrologic parameters should be close to their population values. Nevertheless, an analysis of the sensitivity of the system response, measured in terms of average annual returns, to variations in these parameters, and in the interest rate is desirable.

The effect on the operating policy of different interest rates is examined first.

Next, to determine the effect of incorrect estimates of hydrologic parameters, the question is asked: what is the effect on expected annual returns and on the shortage index of error in estimates of parameters?

In answering this, only the simple case of error in one parameter at a time is considered. The compounded effect of errors in two or more parameters at once is not considered.

4.8.2 Interest Rate

Dynamic programming solutions were obtained for the 48 year historic record using r = 0% and 10%. These are illustrated in Figures 4.8 and 4.9.

Linear policy estimates for these solutions are shown in Figure 4.5, together with the policy for r = 5%. Note that these policies apply only for $s_t \geq 45 \times 10^2$ ac. ft.

FIG. 4.8 DETERMINISTIC SOLUTIONS FOR r=0%, 5%, 10%
HISTORIC DATA (1918-1941)

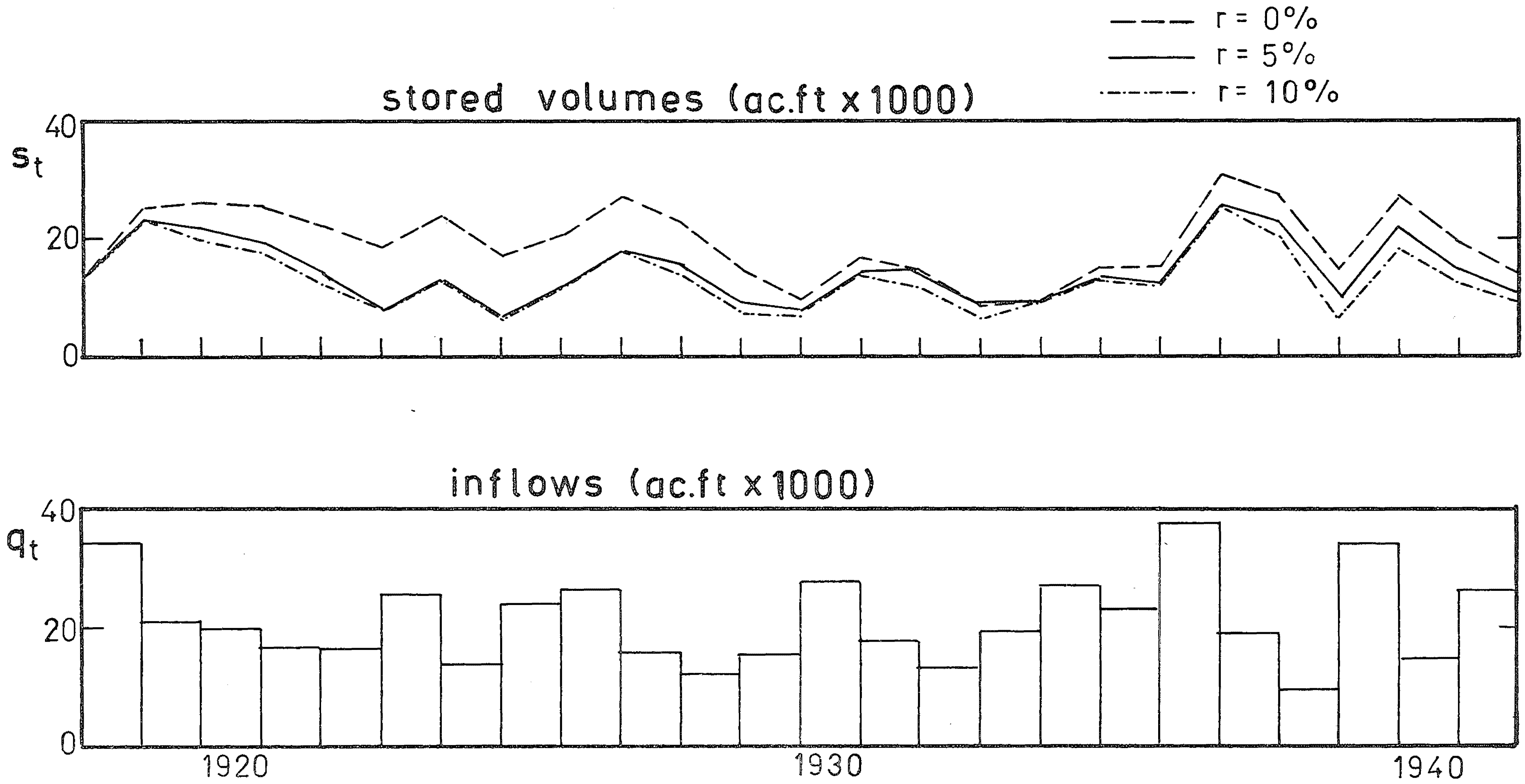
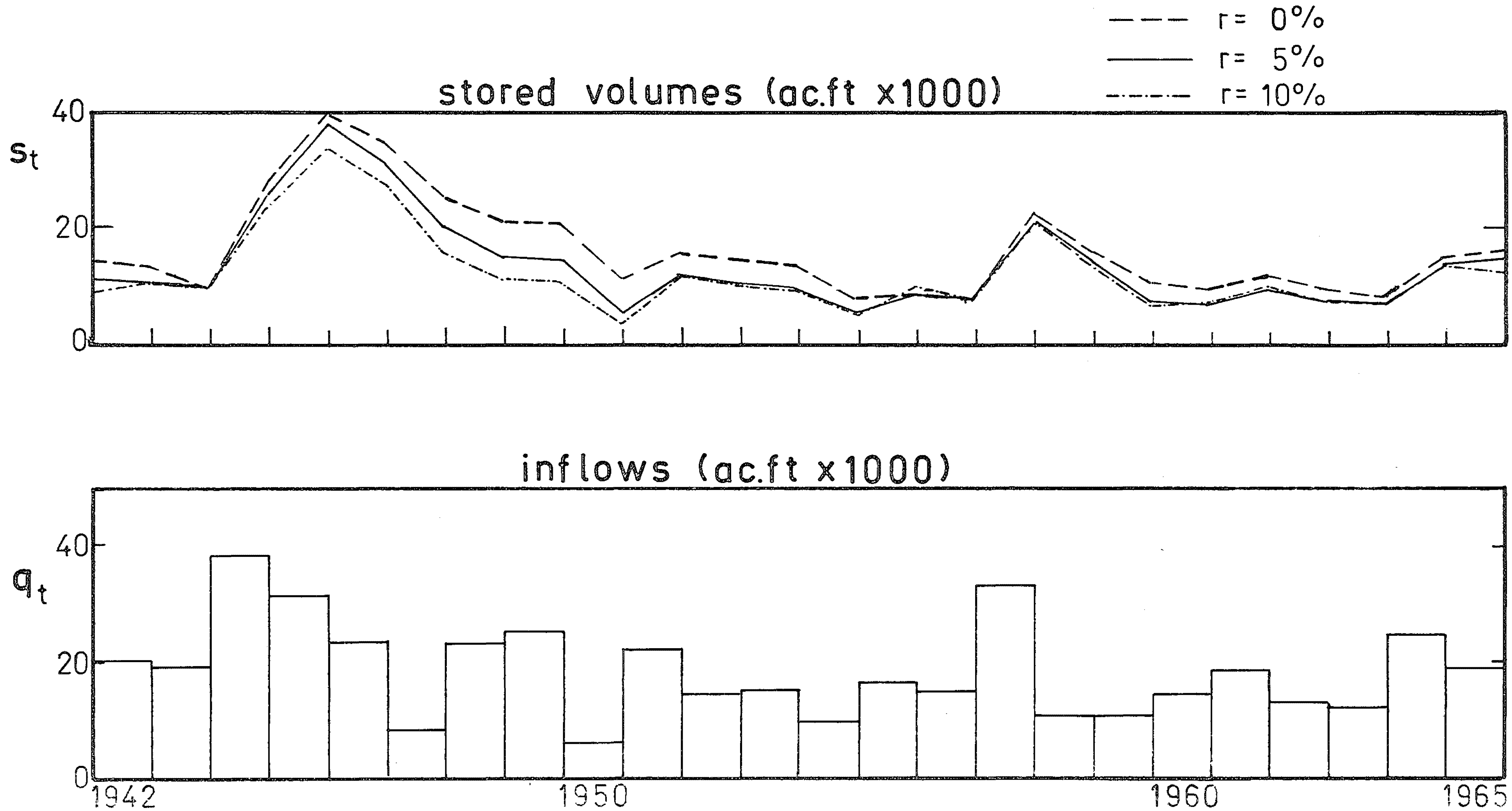


FIG. 4.9 DETERMINISTIC SOLUTIONS FOR r = 0%, 5%, 10%
HISTORIC DATA (1942-1965)



The 0% line is flatter and lower than those for r = 5% and 10%, representing a preference for steadiness in supplies from one year to another. By contrast, the 10% line is steep and is above the others, indicating a preference for using available water immediately, rather than storing it for the future.

Table 4.3 gives the results of using these different policies to operate the storage for five different sequences of inflows. Average annual returns are almost unaffected by variation in the interest rate. However, the shortage indices, and standard deviations of annual returns, increase with the interest rate.

4.8.3 Correlation Coefficient

The autocorrelation coefficient for 48 years historic data has been taken as zero.

To investigate the effect of autocorrelation in the flow data, while using the optimum linear policy, 600 year sets of data were generated from the same sequence of random numbers but with $\rho(1) = 0.0, 0.3, 0.6, 0.8, 0.9$ and 0.95. The other statistics for these data sets were all close to the values for the historic data.

Each of these data sets was routed through the storage using the optimal linear policy, (derived assuming zero autocorrelation), to specify releases.

Results, in terms of average annual returns, standard deviations of annual returns and shortage indices, are given in Figure 4.10.

Figure 4.10 shows that the average annual return decreases and that the shortage index increases as the correlation coefficient increases. These results obtain intuitively; a higher autocorrelation coefficient implies longer runs of wetter than average and drier than average periods, resulting in a higher frequency of emptiness and overflowing at the storage. Annual returns are reduced; their variability and the shortage index increases.

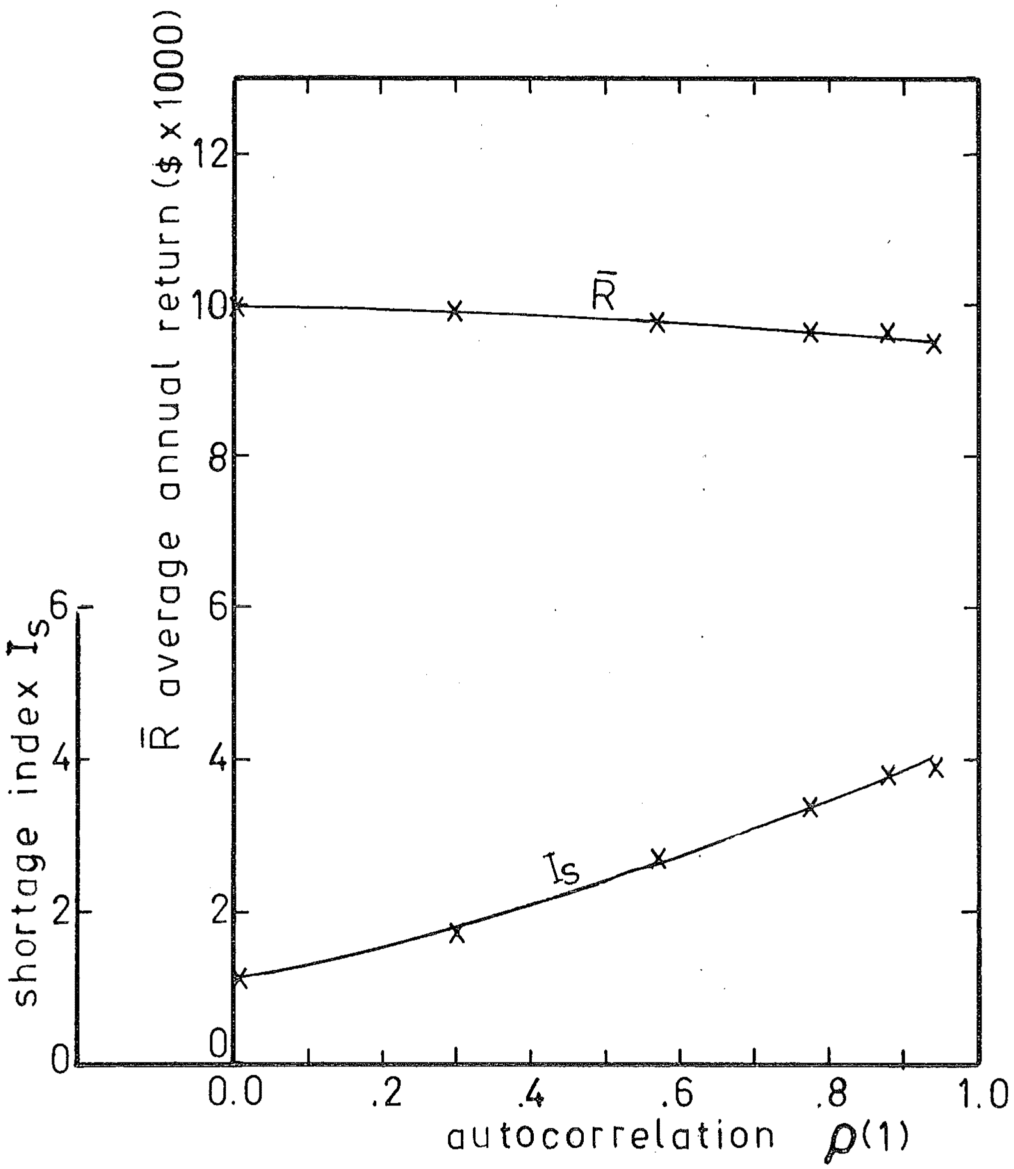
TABLE 4.3

MANORBURN DAM

RESULTS OF DETERMINISTIC SOLUTIONS, r = 0%, 5%, 10%.

Flow Data		Returns (Dollars)						Shortage Indices		
File Name	Mean Annual Flow (ac.ft.)	Average Annual Returns			Std.Devs, Annual Returns			I_s		
		r = 0%	r = 5%	r = 10%	r = 0%	r = 5%	r = 10%	r = 0%	r = 5%	r = 10%
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
MANB1	17600	9480	9470	9450	1090	1320	1420	2.18	2.62	2.52
MANB2	19800	9970	10060	10070	900	1050	1170	1.07	1.34	1.15
MANB3	19700	10140	10020	10190	730	1040	1310	0.63	1.01	1.41
MANB4	18500	9890	9860	9830	800	1100	1190	0.89	1.42	1.38
MANB5	18600	9930	9950	9930	660	770	870	0.61	0.68	0.70
Means for 5	18800	9880	9870	9870	840	1060	1190	1.08	1.41	1.43

FIG. 4.10 SENSITIVITY ANALYSIS-EFFECT OF NEGLECTING SERIAL CORRELATION



4.8.4 Mean Annual Flow

A similar investigation was made assuming that the mean annual flow used in deriving the linear policy was incorrectly estimated. Sets of inflow data 96 years long were generated assuming that the estimated mean was in error by -10%, -5%, 0%, 5% and 10%.

Again, the results in Figure 4.11 of routing these data sets through the storage were predictable. Figure 4.11 shows the shortage index increasing and the average returns decreasing as the true mean flow reduces.

4.8.5 Standard Deviation of Annual Flows

Similar data sets were generated assuming the estimated standard deviation of annual flows was in error by -30%, -15%, 0%, 15% and 30%.

Figure 4.12 illustrates the result from routing these data sets through the storage.

Here the average return increases slightly and the shortage index decreases as the true standard deviation decreases.

4.9 SUMMARY

A rule for releasing water from the Manorburn Dam has been derived using a simulation-optimization-regression approach. With the objective of maximizing the expected present value of annual returns, the rule gives anticipated releases for an irrigation season as a linear function of the volume in storage at the beginning of the season.

The bilinear benefit function which actually applied to the system did not value steadiness in supplies between seasons. For this reason an approximating quadratic benefit function was used in the analysis. This would cause a small upward bias in all the average annual return figures quoted.

The derived real-time policy closely approximates the deterministic solution. In this situation there is little value in knowing future flows; absolute maximum returns were only about 0.7% greater than expected maximum returns for real-time operation.

FIG. 4.11 SENSITIVITY ANALYSIS - EFFECT OF INCORRECTLY ESTIMATING MEAN ANNUAL FLOW

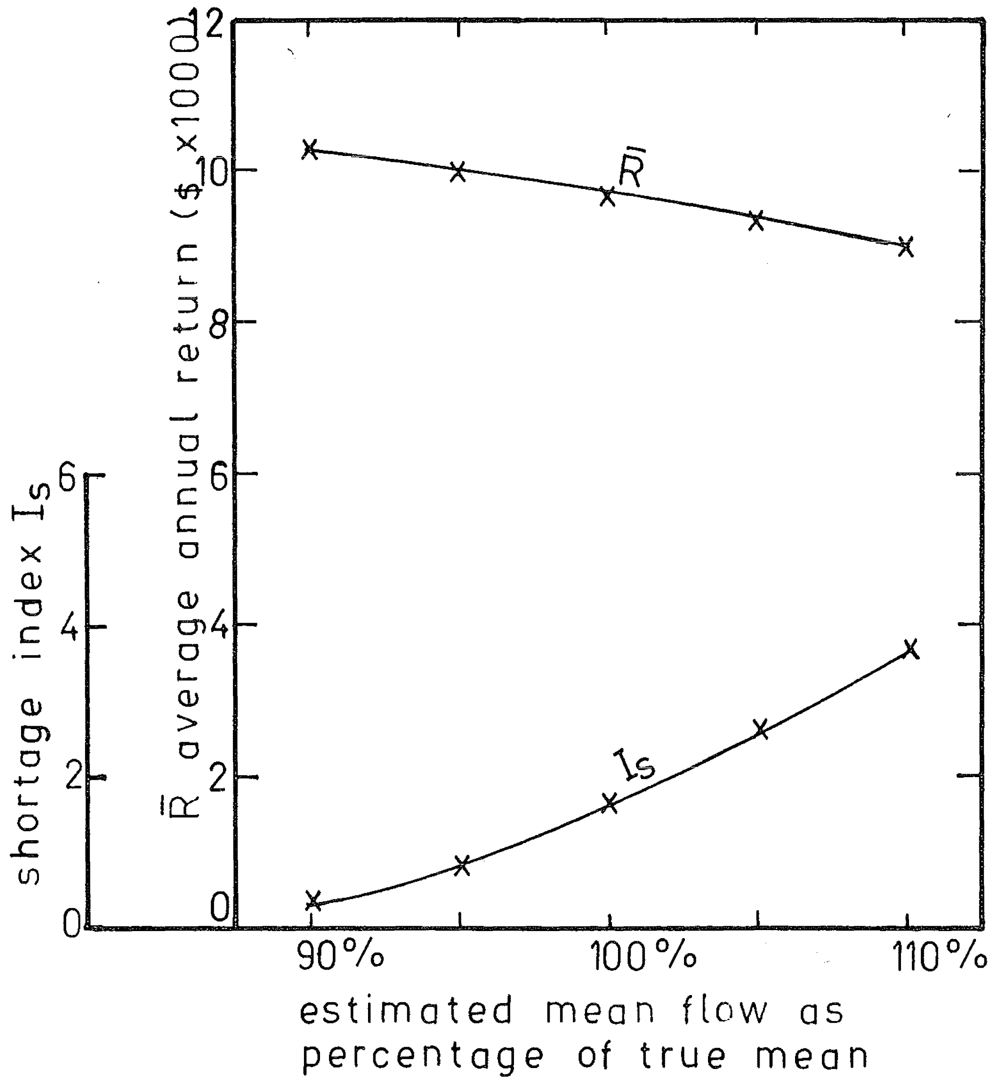
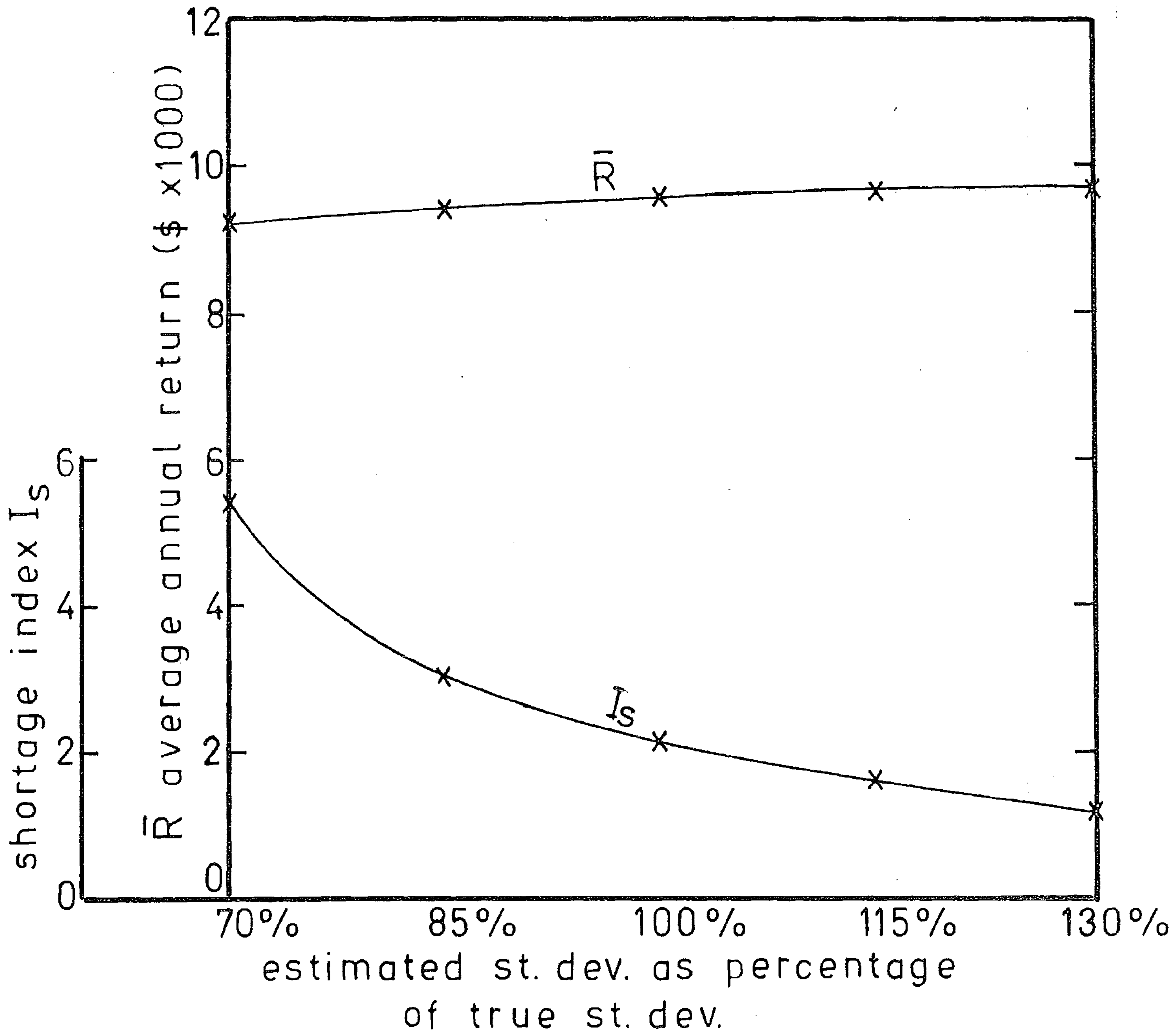


FIG. 4.12 SENSITIVITY ANALYSIS - EFFECT OF INCORRECTLY ESTIMATING STANDARD DEVIATION OF ANNUAL FLOWS



This verifies the intuitive observation that a large storage volume is a good hedge against an uncertain future.

On average, the derived optimal linear policy gave a 0.4% improvement over the present policy in terms of average annual returns.

Nevertheless the present policy may be more desirable because of its ability to make more reliable predictions of seasonal releases.

Average annual returns were not greatly affected by possible serial correlation of annual inflows, or by variations in the annual interest rate. They were, however, quite noticeably changed by possible errors in the mean annual flow. Errors in estimating the standard deviation of the annual flows had little effect on annual returns.

CHAPTER FIVE

OPERATION OF A HYDRO-THERMAL POWER SYSTEM

5.1 INTRODUCTION

This chapter describes a second more complex example of the application of the three step simulation-optimization-regression concept for deriving operating rules. As in Chapter 4, operation with the derived real-time policy is compared to the global optimum obtained from direct dynamic programming solutions for particular sets of inflow data.

The example taken is the Waitaki hydro-electric power system in the South Island, New Zealand.

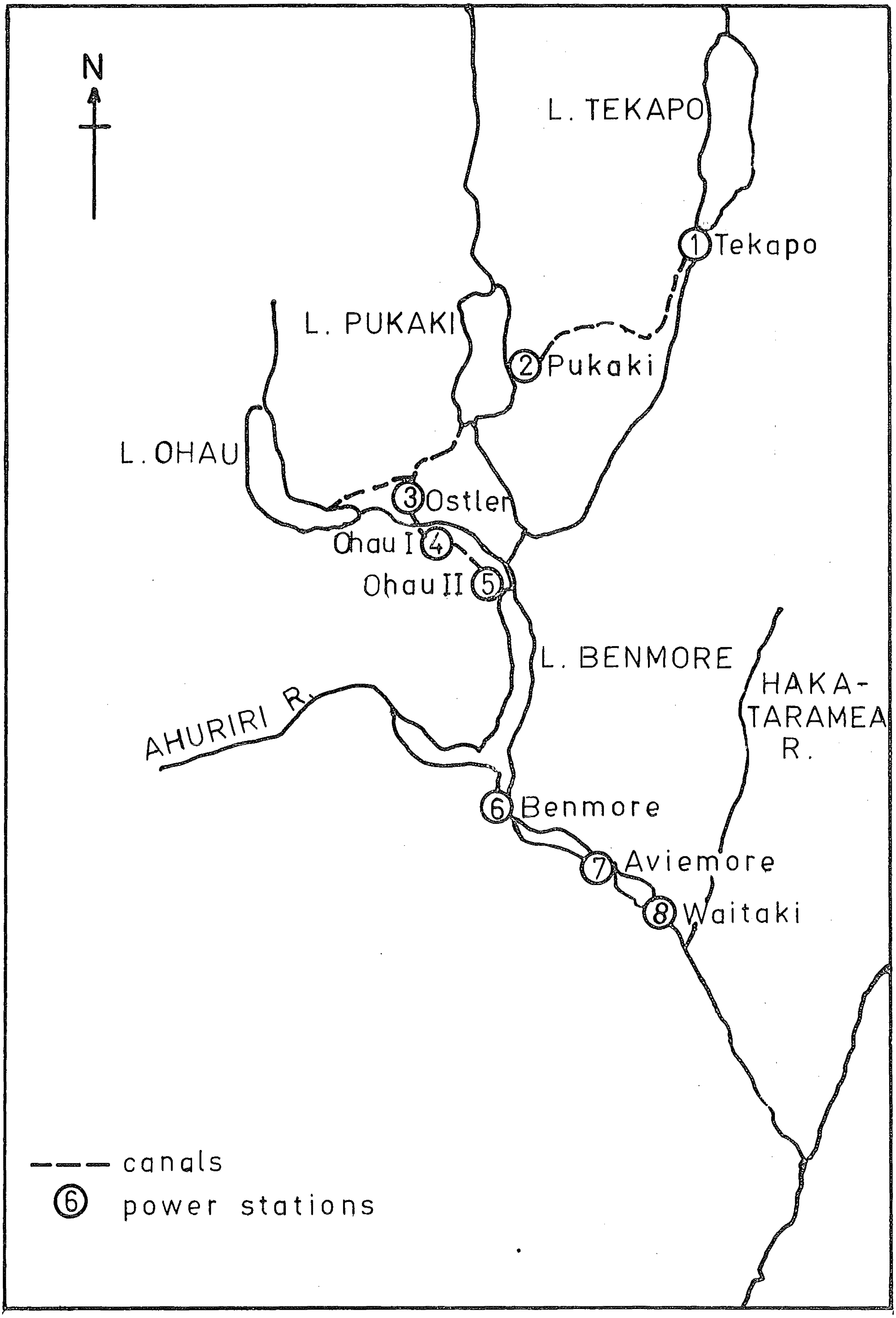
Prior to the development of this system, outflows from three natural lakes, Tekapo, Pukaki and Ohau, combined to form the Waitaki River. Two smaller rivers, the Ahuriri and the Hakataramea enter the Waitaki downstream.

As a fully developed hydro-electric power system, the Waitaki River is shown in Figure 5.1. This system involves a chain of dams, power canals and eight power stations. Part of this system is in operation and the remainder is scheduled for construction.

This hydro-electric system is a segment of the national power system which is in the form of an interconnected grid incorporating both hydro and thermal power generating plants.

In reality, an operating policy for the Waitaki segment cannot be considered in isolation from the rest of the national system. Quantities of power to be generated by the Waitaki system over any time interval will be dependent on many factors, important ones being; the availability of thermal plant, the water storage and the flow situations at hydro plants within and

FIG. 5.1 UPPER WAITAKI POWER DEVELOPMENT - GENERAL PROPOSALS



outside the Waitaki system, the national load, and the transmission line losses between generating and consuming centres.

Furthermore, operation is within an unsteady environment; change is the norm. Power demands, which vary by seasons within a year, generally increase between years. As new hydro and thermal stations are constructed, old thermal plants are relegated to standby roles, and others are scrapped. Fuel costs fluctuate, and with technological improvement the efficiency of new thermal plants increases. In contrast, hydro power plants have long lives; their efficiency is high and is unlikely to be improved in the future. Load centres grow unevenly as population and industry increase more rapidly in some parts of the country than in others. Discoveries of new fuel reserves such as natural gas offer new generating possibilities.

In addition, within any particular hydro-electric system, other purposes such as flood control, recreation and fishing may need to be considered. Hydrologic uncertainty introduces further difficulty.

All these features combine to make the definition of an operating policy for an interconnected hydro-thermal power system a complex problem. Models for solution necessarily require simplifying assumptions and approximations.

The model to be described in this chapter has a number of these, but it is based on a real system, and does take particular account of hydrologic uncertainty.

5.2 DESCRIPTION OF THE WAITAKI HYDRO-ELECTRIC SYSTEM

Within the fully developed system which is illustrated in Figure 5.1, eight power stations are proposed. Details of these are given in Table 5.1. Active storage volumes of the dams are given in Table 5.2. Outflows from Lake Ohau are uncontrolled, but the dams at the outlets of Tekapo and Pukaki provide active storage volumes that can hold wet season runoff for power generation in dry seasons.

TABLE 5.1

POWER STATION DETAILS

(1) Station i	(2) Installed Capacity MW	(3) Max. Head (Gross)	(4) Efficiency S-station M-machine	(5) Power Cusec/MW	(6) Energy P_i Gwh/1000 CSD	(7) Full Load Flow 1000 CSD / month	(8) Ann. Output at Mean Ann. Flow Gwh
1	25	104	0.8 S	160	.149	136	161
2	160	484	.89 M	28	.860	136	920
3	264	190	0.87 S	714	.340	554	1320
4	224	170	0.87 S	80	.300	554	1180
5	200	150	0.87 S	90	.267	554	1050
6	540	305	.92 M	42.5	.565	700	2040
7	220	123	.92 M	111.1	.216	730	960
8	105	70	.88 M	194	.124	638	535

TABLE 5.2

STORAGE LAKES

Lake	Operating Range (feet)	Storage CSD x 1000
Tekapo (A)	2335 - 2310	272
Pukaki (B)	1741 - 1696	760
Benmore (C)	1181.5 - 1181	5
Aviemore	876 - 874	7
Waitaki	753 - 746	5

Within the operating ranges for Lakes Benmore, Aviemore and Waitaki, only relatively small volumes of storage are available, and effectively the generating plant below each of these lakes are constant head.

Lakes Aviemore and Waitaki are relatively small in area and the volumes obtained by drawing these lakes down are small compared to the volumes available at Pukaki and Tekapo. Benmore, however, when full, has a surface area of 19,000 acres, compared with 23,200 acres for Tekapo (full) and 42,100 acres for Pukaki (full). If an operating range of say 20 feet were allowed for Benmore instead of the 0.5 feet indicated in Table 5.2, an additional active storage volume of 179,000 CSD would be available. Thus in terms of month-by-month operation, there are three storages in this system.

Under the conditions of a dry winter with the other two large storages drawn down, this additional volume at Benmore could be most useful. However, as indicated in Table 5.2, the New Zealand Electricity Department does not intend to use this volume.

For this reason, and also because the conjunctive operation problem for a three lake system implies a three state variable dynamic programming problem, which is computationally very demanding to solve directly, the extra volume available at Benmore will not be considered here. Instead, the conjunctive operation problem for the volumes available at Tekapo and Pukaki, a two state variable problem, will be solved.

Further, if Benmore is held at a fixed level, the only variable head power station in the system, No.1 in Figure 5.1, is at the outlet of Lake Tekapo.

After this station, controlled releases from Tekapo are diverted through a power canal to Lake Pukaki, passing through one further station on the way. Controlled releases from Lake Pukaki merge with the outflows from Lake Ohau and pass through power stations 3, 4 and 5 before entering Lake Benmore. The Ahuriri River enters Lake Benmore directly, as do spills from Tekapo and Pukaki. Releases from Lake Benmore generate power at stations 4, 5 and 6.

With three storages, the hydro system is represented by the flow chart in Figure 5.2. Inflows to the system are represented by q_1, q_2, q_3 and q_4, which correspond to the flow recording stations 1, 2, 3 and 4 listed in Table 2.1. Statistical analyses of the monthly flow records for these stations have been described in Chapter 2. A multivariate model to represent monthly flows at these stations is described in Chapter 3.

For continuity, overflows from storages, and seepage and evaporation losses are also shown.

The long range operating problem for the system outlined in Figure 5.2 is to allocate water between current and future use. The long range problem is considered in terms of month-by-month decision making. (In this chapter, a month is defined as a standard time unit of 30.44 days; this is a slight change from the calendar definition used in Chapters 2 and 3.)

There is also an interesting short term problem which is not examined here - how best to meet the instantaneous power demands during a day.

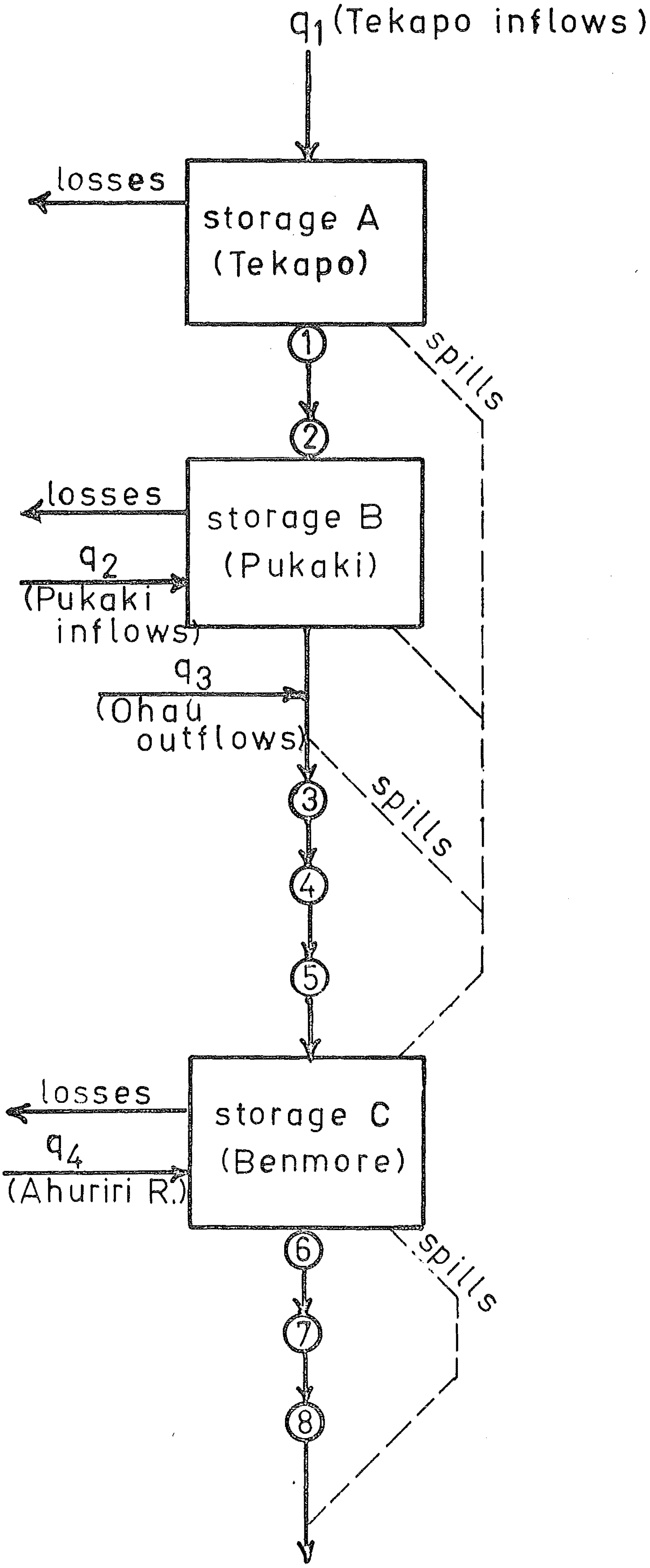
5.3 MODEL OF HYDRO THERMAL POWER SYSTEM

5.3.1 Introduction

When other possible purposes of the Waitaki system, such as recreation and irrigation, are neglected, and the active storage available at Benmore is taken as zero, there remains the long range operating problem of allocating water in Lakes Tekapo and Pukaki (storages A and B in Figure 5.2) between current and future use.

If the Waitaki system was independent of other generating plant, and was to be operated by a concern which supplied power to consumers, a contract would specify levels of "firm" energy supplies to consumers for specified time periods, with penalties for failure to deliver the full "firm" supplies. In addition, "non-firm" or "dump" power, generated in times of good inflows with water which would otherwise be spilled to waste, might be sold at cheap rates. In this situation the objective in operation of maximizing the present value

FIG. 5.2 FLOW CHART FOR PROPOSED SYSTEM



of revenue less operation, maintenance and repair costs would be reasonable; this is analogous to the objective in operating the Manorburn Irrigation Dam, which was described in Chapter 4.

5.3.2 Generating Costs

In reality, the system is part of a national grid, and if a power load is not met by generation within the Waitaki, it must be generated elsewhere, since power shortages are not admissible. Alternative sources are thermal plants, and other hydro plants outside the Waitaki.

In a steady system, all the plant has been constructed and the load is constant from one year to another. In terms of real-time month-by-month operation, capital costs are irrelevant, the aim should be to minimize the present value of generating costs.

Generating costs are subdivided as; (a) fuel and operation, maintenance and repair (O.M.R.) costs for thermal plants and (b) O.M.R. costs for hydro plants. In comparison with thermal plants, generating costs for hydro plants are low and are largely independent of the quantities of power generated by a plant.

In contrast, thermal generating costs are much greater and are expected to be related to quantities of power generated. Costs are expected to rise rapidly as the thermal energy output increases because;

(1) Increased quantities of fuel are required,

(2) Plants such as gas turbines, designed specifically for peak load running, are inefficient and expensive to use for high load factor base load operation,

(3) Old standby plants brought into operation to meet increased loads are expensive to operate in comparison to new more efficient thermal plants.

Thus thermal generating costs are expected to increase rapidly as the rate of energy output increases.

On a monthly basis, the cost of thermal energy can be approximated by a polynomial function of the type -

$$C_n = c_o + c_1 g_n + c_2 g_n^2 + c_3 g_n^3 + \ldots\ldots \qquad (5\text{-}1)$$

where C_n is the thermal cost for the nth month,

$c_o, c_1, c_2 \ldots$ are constants,

g_n is the quantity of thermal energy generated in the nth month.

Equation (5-1) is an approximation to an unknown cost function, which probably has discrete steps as more costly plants come into operation to meet increased loads. Quadratic and linear cost functions have been used in this type of study (Little, 1955).

In this study, quadratic costs will be assumed. Taking c_0 and $c_1 = 0$,

$$C_n = c_2 g_n^2 \qquad (5\text{-}2)$$

Estimated on 1968 costs, $c_2 = 0.0005$, where g_n is in Gwh (Kwh x 1000), and C_n is in \$ x 1000. Thermal generating costs are the only costs considered here. Thus, provided c_2 is positive, its numerical value does not affect an operating policy. Of course the outcome from using a particular policy, measured in terms of thermal costs, is dependent on c_2.

5.3.3 Energy Load

As an approximation in setting up a model to study the operation of the Waitaki hydro-electric power stations, alternative hydro generation available in other parts of the country will be neglected. The Waitaki stations, in conjunction with thermal power stations, will be required to generate given monthly energy loads.

Under average conditions, the annual energy output from the Waitaki stations is expected to be 8566 Gwh/year. Assuming that the expected annual thermal load is 10% of the expected hydro load, the combined annual load is 9423 Gwh.

The average monthly load pattern for the national system is given in column 2 of Table 5.3. This pattern is in terms of an arbitrary power unit; the maximum occurs in month 5. Monthly energy loads K_j in column 3 of Table 5.3 are obtained by dividing out the annual load in accordance with this

TABLE 5.3

MONTHLY POWER LOADS

(1) Month j	(2) % of 1 P.U.	(3) Load (Gwh/month) (1 P.U. = 982 Gwh/month) K_j
1 (Jan)	72	704
2	79	774
3	89	873
4	96	940
5	100	982
6	93	912
7	82	803
8	75	745
9	72	704
10	65	637
11	67	657
12	72	704

pattern. It is assumed that this set of loads will be the same for each year of operation.

5.3.4 Objective

The objective in operation over a period of N months is taken as minimizing the present value of thermal generating costs.

Given the power load in the nth month as K_n, let H_n be the quantity of hydro power generated. Thus the thermal power requirement is;

$$g_n = K_n - H_n \tag{5-3}$$

$H_n > K_n$ corresponds to the generation of dump power with water which would otherwise be spilled. In this case $g_n = 0$. Dump power is not valued in this study.

C_n, the cost of generating g_n thermal units, is given by (5-2). The present value of the cost of operating the system over N periods is given by

$$PV = \sum_{n=1}^{N} \frac{1}{(1+r)^n} C_n \tag{5-4}$$

where r is the monthly discount rate. r is taken to give an equivalent annual discount rate of 8%.

The objective assumed in operating the combined hydro-thermal power system is the minimization of the present value of generating costs, given by (5-4).

5.3.5 Seepage and Evaporation Losses

Losses from the storages are mainly in the form of seepage and evaporation.

Seepage is a function of the geological conditions at the storage and in the power canals. Little information on the evaluation of seepage losses is available. Probably they are low, because the fine glacial silt carried in the water tends to seal porous soils. In this study, seepage losses are neglected.

Net evaporation from a storage surface is dependent on the time of year. The loss L in the nth month is approximated by the expression

$$L = A_n . (E_n p - R_n), \tag{5-5}$$

where A_n is the surface area in the nth month, computed as the area corresponding to the average of initial and final volumes in storage in the nth month,

E_n is the expected pan evaporation in the nth month,

p is the open water reduction factor $\simeq 0.8$,

R_n is the expected rainfall in the nth month.

Areas of the storage surfaces were computed from the appropriate elevation/area and elevation/volume curves. Pan evaporation figures were mean recordings taken from a soil conservation station within the region. Average monthly rainfall figures were from records made at Lake Tekapo. These are taken as representative figures which are applied over the whole region.

Net evaporation losses are relatively small; peak values in summer months are about 1% of the active storage volumes.

5.3.6 The Functional Equation

The general functional equation of dynamic programming was described in Chapter 1. This equation is rewritten for the specific system shown in Figure 5.2, but not taking account of storage C, the volume available at Benmore. Define;

S_A as the active storage volume at Tekapo,

S_B the active storage volume at Pukaki,

D_A the maximum discharge through turbines at Tekapo,

D_B the maximum useful release from Pukaki,

s_A, s_B the volumes stored in A and B at the beginning of the nth month,

s'_A, s'_B the volumes stored in A and B at the end of the nth month,

d_A, d_B the releases from A and B during the nth month,

q_1, q_2, q_3, q_4, the inflows to the system during the nth month,

C_n (s_A, s_B, d_A, d_B), the cost of thermal generation in the nth month, as calculated from (5-2) and (5-3).

L_A (s_A, s'_A), the loss from A during the nth month,

L_B (s_B, s'_B), the loss from B during the nth month,

z_A, z_B, the volume of spills from A and B during the nth month,

f_n (s_A, s_B) the present value of thermal costs for n future periods of operation starting with volumes s_A and s_B in storage and following an optimal policy throughout. It follows that f_{n-1} (s'_A, s'_B) is the present value of thermal costs for (n-1) future periods of operation.

Thus,

$$f_n(s_A, s_B) = \min \left(C_n(s_A, s_B, d_A, d_B) + \frac{1}{1+r} f_{n-1}(s'_A, s'_B) \right) \qquad (5\text{-}6)$$

subject to the following constraints;

$$0 \leqslant d_A \leqslant \min (D_A, (s_A + q_a - L_A(s_A, s'_A))) \qquad (5\text{-}7)$$

$$s'_A = \min (S_A, (s_A + q_1 - d_A - L_A(s_A, s'_A))) \qquad (5\text{-}8)$$

$$z_A = \max (0, (s_A + q_1 - d_A - L_A(s_A, s'_A) - s'_A))$$

$$0 \leqslant d_B \leqslant \min (D_B, (s_B + q_2 + d_A - L_B(s_B, s'_B))) \qquad (5\text{-}9)$$

$$s'_B = \min (S_B, (s_B + q_2 + d_A - d_B - L_B(s_B, s'_B))) \qquad (5\text{-}10)$$

$$z_B = \max (0, (s_B + q_2 + d_A - d_B - L_B(s_B, s'_B) - s'_B))$$

5.3.7 Evaluation of Losses

Evaporation losses are evaluated in terms of the mean volume in storage during the nth interval. This causes s'_A and s'_B to appear in the right hand sides of (5-8) and (5-10) respectively. Explicit solutions for s'_A and s'_B are not undertaken. Since losses are small (1% or less of stored volumes), initial estimates of s'_A and s'_B are made by neglecting the loss terms in

(5-8) and (5-10). With these estimates, mean stored volumes and hence surface areas and losses are found. With these losses, revised estimates of s'_A and s'_B are obtained from (5-8) and (5-10) respectively.

5.3.8 Estimation of Energy Output

Define h_i as the energy output from the ith station during a particular month, and P_i as the energy output constant in Gwh/1000 CSD. P_i values corresponding to full head are in the 6th column of Table 5.1.

At the outlet from storage A, station 1 has a variable head, depending on the stored volume. To estimate energy output during a particular month, the head corresponding to the mean stored volume is used.

Thus the energy output from station 1 is

$$h_1 = \frac{\bar{x}}{104} \cdot P_1 \cdot d_A$$

where $\bar{x}$ is the head corresponding to the mean stored volume.

P_1 is the energy constant in Gwh/1000 CSD corresponding to a full head of 104 feet.

d_A is release through the turbines in CSD x 1000.

In this model, the remaining stations are fixed head.

Station 2,

$$h_2 = P_2 \cdot d_A$$

Stations 3, 4, 5,

$$h_i = P_i d' \quad (i = 3, 4, 5)$$

subject to

$$d' = \min\left((q_3 + d_B),\ D' \right),$$

where d' is the discharge through 3, 4, and 5,

and D' is the maximum discharge through 3, 4 and 5.

Stations 6, 7 and 8,

$$h_i = P_i d'' \quad (i = 6, 7, 8)$$

subject to

$$d'' = \min\left((q_3 + q_4 + d_B + z_A + z_B),\ D''\right)$$

where d'' is the discharge through 6, 7 and 8,

and D'' is the maximum discharge through 6, 7 and 8.

Maximum useful discharges are shown as "full load flow" in column 7 of Table 5.1.

5.4 SOLUTION OF THE MODEL

5.4.1 Computer Program

To numerically solve the functional equation (5-6) a computer program was written in FORTRAN IV for implementation on the IBM 360/44 installation at the Canterbury University Computer Centre.

Discrete levels of storage volumes and releases specified for the storages A and B were (in CSD x 1000)

Storage	Step Size	Maxm. Vol. (i.e. active storage)	No. possible storage levels	Max. Rel.	No. possible release levels
A	34	272	9	136	5
B	40	760	20	520	14

The 5-40 year sets of inflow data described in Chapter 3 were used to obtain five corresponding sets of optimal releases and storage levels.

An extra 12 months data was added at the end of each 40 year flow sequence. Carrying the optimization through this extra 12 months removed most of the influence of the terminal state (at the end of the 41st year) from the 40 years of deterministic solution. Thus the 40 years of solution is "steady-state" - expected costs are constant for each of the 40 years. Results for the 41st year were discarded.

As programmed by the author, the time taken by the deterministic optimization phase of the program for a 41 year sequence of flow data was 97 minutes. This execution time is linearly dependent on the number of years covered and represents approximately 12 seconds per month. The execution time is a function of the number of storage and release steps.

The particular combination of step sizes and length of record chosen here gave a suitable execution time, whilst the storage requirements of the program remained within the capabilities of the installation. No attempt was made to reduce this time requirement, although several courses of action were open:

(1) increase the step sizes,

(2) rewrite, in more efficient ASSEMBLER language, that part of the FORTRAN program, some 32 statements, within which most of the execution time was spent,

(3) reduce the number of seasons considered within a year to say 6 or 4, giving a bimonthly or a quarterly operating problem.

Further details of the program and an analysis of storage requirements for solving dynamic programming problems are given in Appendix I.

5.4.2 Results

The minimum annual cost for operation with each data set is given in column 2 of Table 5.4. Standard deviations of annual costs are in column 2 of Table 5.5.

TABLE 5.4

AVERAGE ANNUAL COSTS
(Thousands of dollars)

Data Set	D.P. soln.	Real-time operation with policy from data set no.				
		1	2	3	4	5
(1)	(2)	(3)	(4)	(5)	(6)	(7)
1	187	-	228	225	224	228
2	98	143	-	131	136	136
3	123	159	141	-	152	151
4	131	165	164	161	-	163
5	123	158	151	148	150	-

TABLE 5.5

ST. DEVS. OF ANNUAL COSTS
(Thousands of dollars)

Data Set	D.P. soln.	Real-time operation with policy from data set no.				
		1	2	3	4	5
(1)	(2)	(3)	(4)	(5)	(6)	(7)
1	129	-	177	175	166	181
2	95	93	-	113	104	124
3	86	76	107	-	88	101
4	83	82	118	111	-	111
5	86	78	111	104	94	-

To illustrate the results of dynamic programming deterministic policy solutions, a four year period of storage behaviour analysis is plotted in Figures 5.3 and 5.5

The full line in Figure 5.3 shows optimal storage volumes for A and B through years 10 and 11 of data set 5; similarly, optimal volumes through years 12 and 13 are in Figure 5.5

Dashed lines in Figure 5.4 show the inflow volumes q_1, q_2 and $(q_3 + q_4)$ for years 10 and 11, and inflows for years 12 and 13 are in Figure 5.6. For comparison the mean inflows for all 40 years of the data set are plotted with full lines in Figures 5.4 and 5.6.

Monthly thermal generating costs for these four years of deterministic solution are the full lines in Figure 5.7. Annual totals for each of the years are in the first row of Table 5.6.

TABLE 5.6

ANNUAL COSTS FOR SAMPLE PERIOD FROM DATA SET 5
(Thousands of dollars)

	Year				Average for 40 yrs of data set 5 (as in Table 5.4)
	10	11	12	13	
Deterministic soln.	18	151	216	294	123
Real-time policy (derived from data set 4)	51	152	130	524	214

FIG. 5.3 STORAGE BEHAVIOUR ANALYSES FOR YEARS 10,11

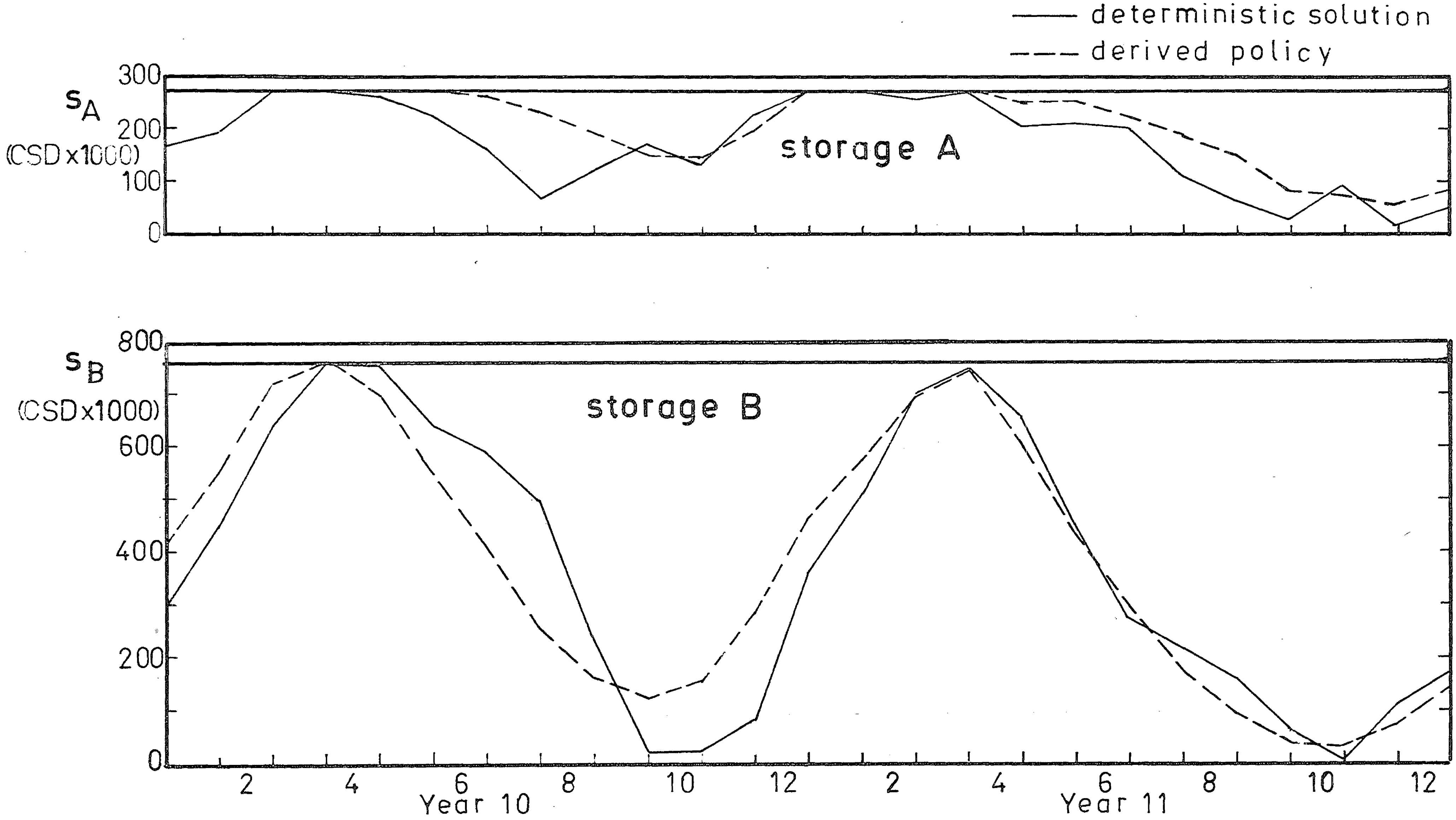


FIG. 5.4 INFLOWS FOR YEARS 10 ,11 (CSD x1000)

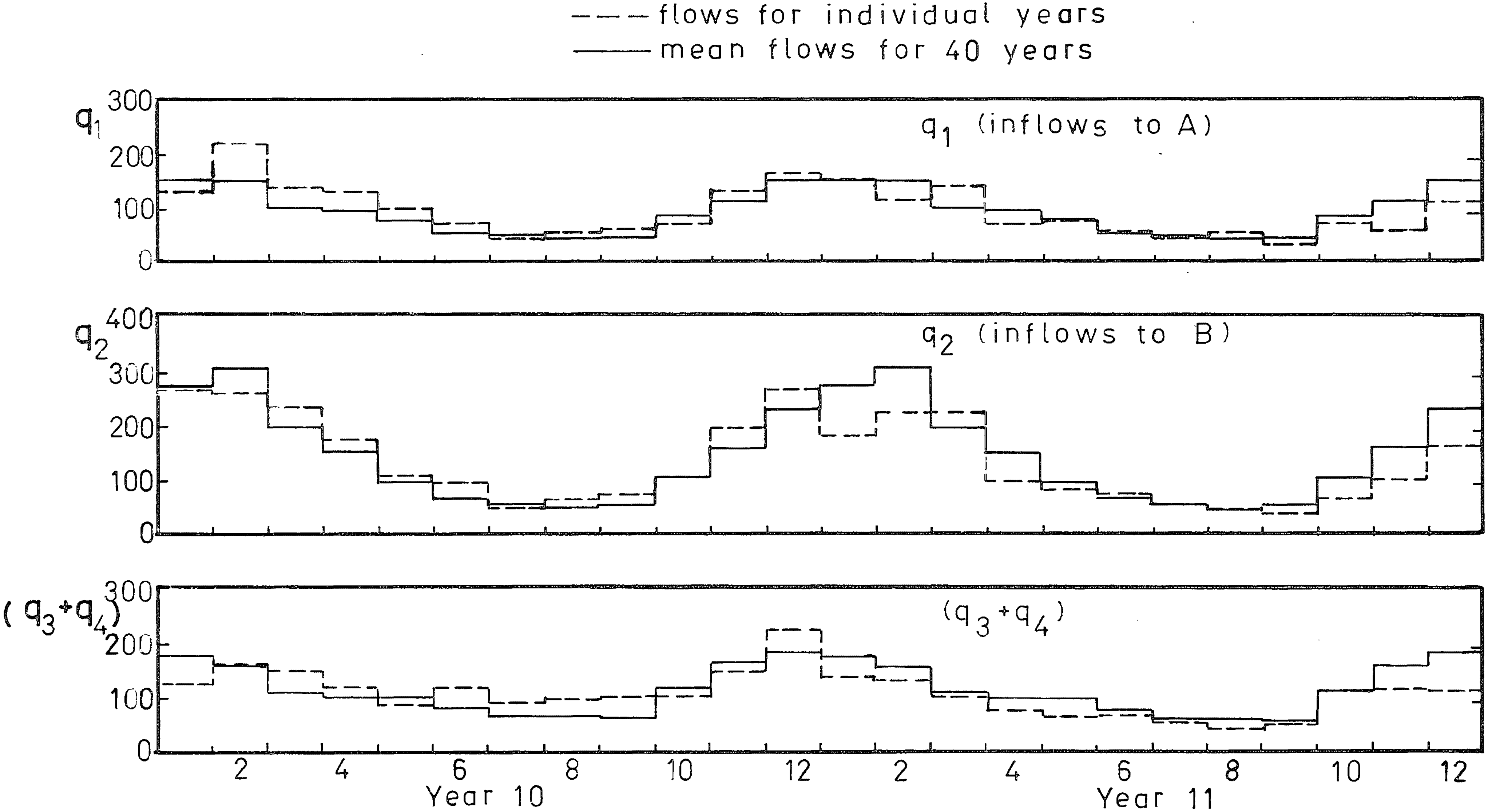


FIG. 5.5 STORAGE BEHAVIOUR ANALYSES FOR YEARS 12,13

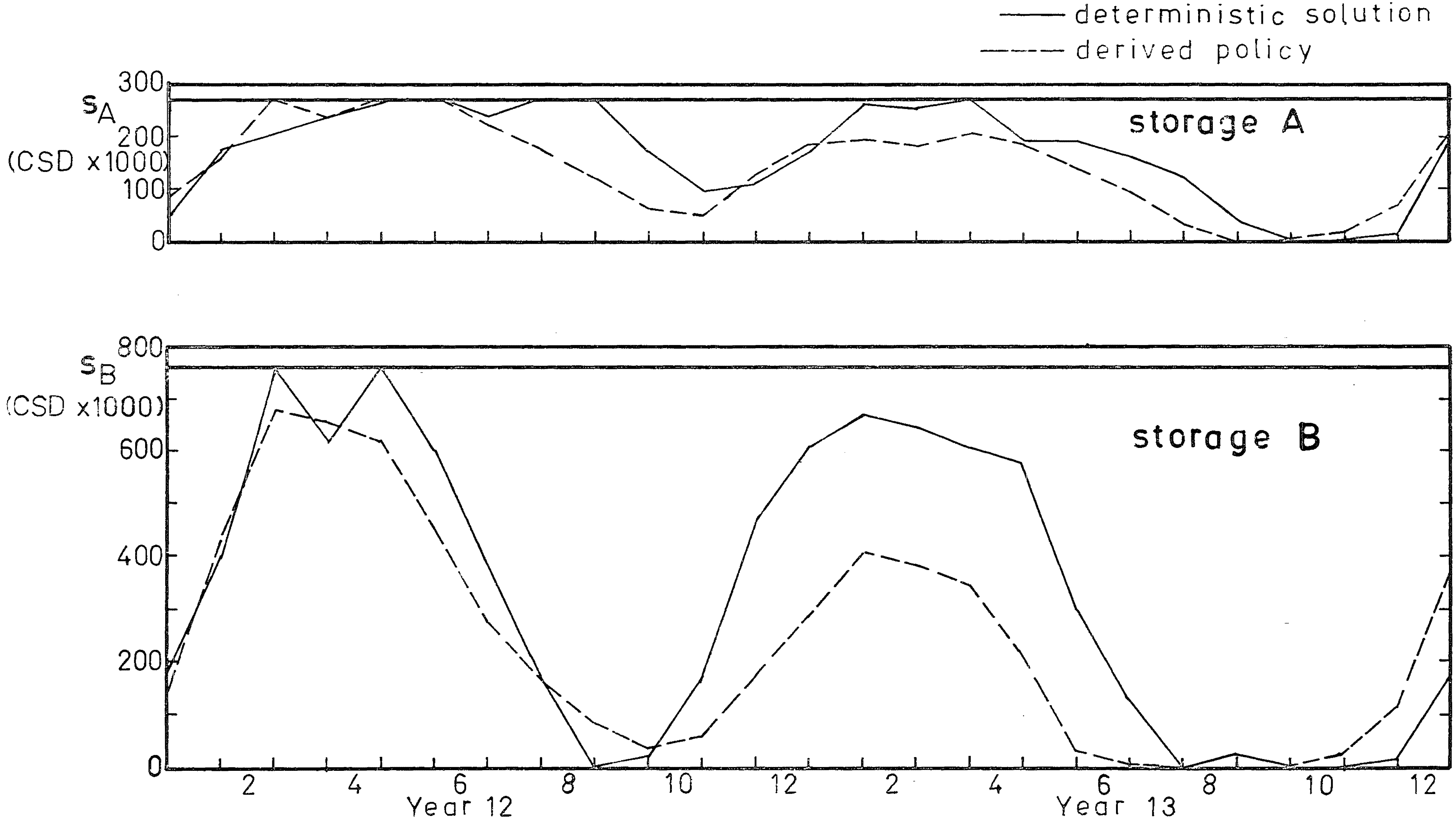


FIG. 5.6 INFLOWS FOR YEARS 12,13 (CSD x1000)

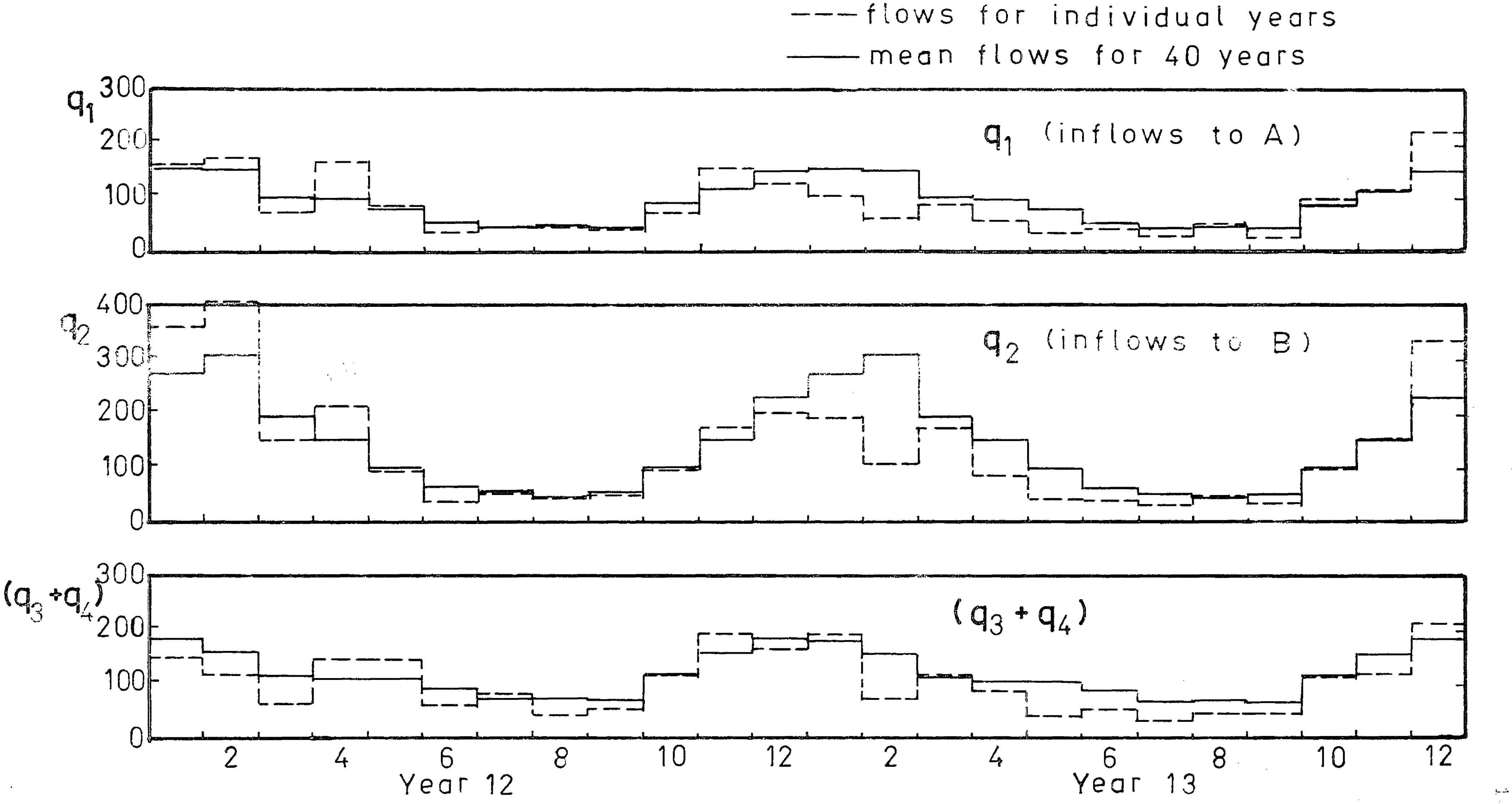


FIG. 5.7 MONTHLY THERMAL COSTS FOR YEARS 10,11,12,13 ($ x1000)

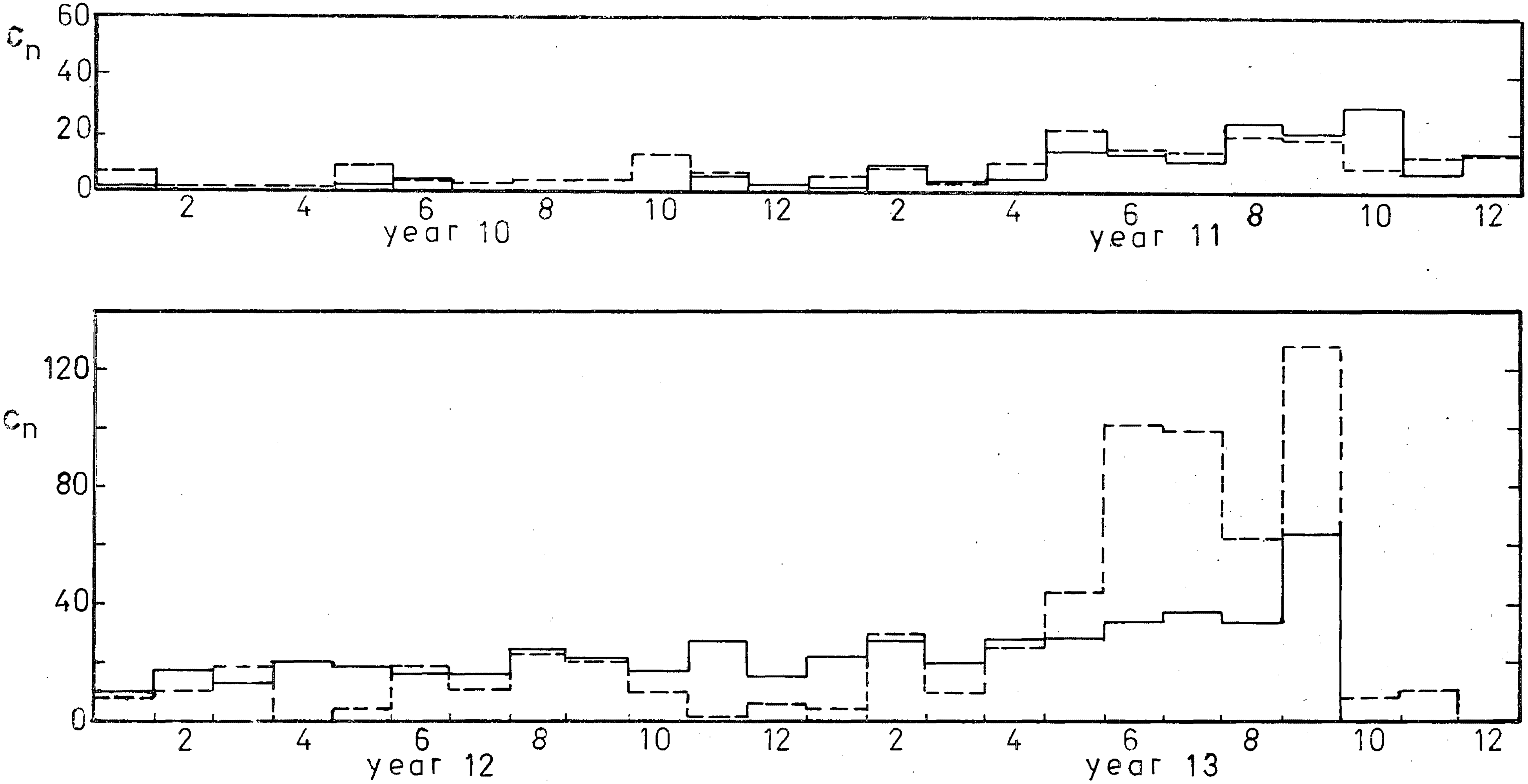


Figure 5.4 shows that inflows in year 10 were frequently above average. (Note the effect of the interstation correlations; flows high at one station imply high flows at the other stations). In this year water was plentiful and the thermal costs (in Figure 5.7) were very low. Both storages overflowed for a period and neither emptied (Figure 5.3).

In year 11, q_2 and $(q_3 + q_4)$ tended to be below average (Figure 5.4), and Figure 5.3 shows the storages drawn well down in months 9 and 10. Larger thermal costs were incurred in this year.

An interesting feature of the deterministic solution is the way it handles a dry spell. Figure 5.6 shows the period from month 12 of year 12 to month 9 of year 13 as being much drier than average. It is especially dry in the first two months of year 13.

Because the deterministic solution effectively "knows in advance" that a dry spell is approaching, it is able to take action to moderate the effect. It does this from the beginning of month 4 of year 12 by incurring moderate thermal costs and building up stores in A particularly. By the end of this year the storages are relatively high (Figure 5.5).

These supplies are rationed optimally through to the end of month 9 of year 13 when both storages become empty. Steady costs are incurred through this period (Figure 5.7). From month 10, with a low load and inflows picking up, costs become zero. For year 13, thermal costs total \$294,000, compared with the average of \$123,000.

The important point here is that a dry spell was anticipated more than eight months in advance and that action was taken by incurring moderate thermal costs over a period of 18 months. In this way extreme hydro power shortages and their associated high costs are reduced substantially.

Many examples of this type of behaviour could be extracted from the optimal data. That given here serves as a good example. The problem is now to use the optimal data to derive a real-time operating policy.

5.4.3 Derivation of the Real-Time Policy

For each 40 year solution, multiple regression was used to determine functions of the type

$$\left.\begin{aligned} d_A &= f\,(s_A,\ s_B), \\ \text{and} \quad d_B &= g\,(s_A,\ s_B), \end{aligned}\right\} \qquad (5\text{-}11)$$

Regression results from fitting the functions

$$\left.\begin{aligned} d_A &= a_0 + a_1\ s_A + a_2\ s_B, \\ \text{and} \quad d_B &= b_0 + b_1\ s_A + b_2\ s_B, \end{aligned}\right\} \qquad (5\text{-}12)$$

are given in Appendix II. For each of the five data sets, 24 policy functions were obtained, one per storage for each month of the year.

Over half of the F values given in Appendix II are significant at the 95% level; implying that over half the multiple correlation coefficients are significant. The explanation of d_A and d_B provided by these functions is related to the seasonal nature of the storage behaviour analyses.

Of the non-significant F values, many occur for the functions pertaining to months 10, 11, 12 and 1. These are the months of late spring, and summer, when water is usually plentiful (Figure 2.1). At this time of year storages are filling after the winter drawdown and power demands are at their lowest and can mostly be met with hydro generation. Except for dry spells, these months are non-critical.

Releases in these months will depend on the volume of flows to come during the rest of the summer and through the fall and winter. If these flows are below average, it is already time to start conserving supplies by using moderate amounts of thermal generation. If these flows are above average, little thought needs to be given to the future and hydrogeneration can be used almost exclusively. Not having information about future inflows, a real-time policy for these months can do no better than to assume average inflows and specify an average release. As the regression analysis indicated, releases for these months are largely independent of the currently stored volumes.

Months 2 and 3, in the late summer and early autumn, correspond to the time when storages, if not already full, should be filling. After month 2, inflows tend to taper off and the power load increases with the onset of winter.

The winter drawdown starts around month 4, with the heaviest power loads for the year in months 4 and 5. About half the F values for months 2, 3 and 4 are significant at the 95% level.

Of the F values for months 5, 6, 7, 8 and 9, 82% are significant. Most of the releases for these months are dependent on the volumes in the storages. As the winter drawdown continues through months 7, 8 and 9, stored water may run out. Supplies remaining at the end of month 9 may be used in months 10 and 11 if spring inflows are below average. Of the total period covered in the deterministic optimization, storage A was emptied 1% of the time and storage B 6% of the time. All of the times that A was empty and 93% of the times B was empty were during months 7, 8, 9, 10 and 11.

In summary, a real-time policy as derived here consists of linear functions of the type (5-12) for all months except 10, 11, 12 and 1, when mean releases are specified.

5.5 REAL-TIME OPERATION

Each of the five sets of policies defined in Appendix II are sample estimates of a real-time policy derived from population data. Use of this population policy over a period of years gives the "expected", rather than the "absolute" average annual cost which the deterministic solution provides.

Given these linear policy estimates in Appendix II, the following questions arise:

1. How variable are these policy estimates - do 40 years of deterministic solution enable a good estimate to be made of the population policy?

2. How do these real-time policy functions work out in actual operation - does the behaviour of the storages operated with a real-time policy resemble their behaviour under the deterministic solution for a given set of inflows - how do expected annual costs compare with absolute annual costs?
3. Are flow forecasts one or more months ahead likely to reduce expected annual costs?
4. How does operation with a derived real-time policy compare with the planned operating procedure of the New Zealand Electricity Department? This question is only valid if one is satisfied that the two storage model used here closely approximates the real system.

The implications of these questions occupy the remainder of this chapter.

5.5.1 Variability of Policy Estimates

From each one of the five data sets, an estimate of a policy for real-time operation has been obtained.

To find how these policy estimates varied between themselves, and how they compared with deterministic solutions, each policy was used to operate the storage behaviour model.

One data set has been used to derive each policy. With each policy, the remaining four data sets were routed through the model to simulate real-time operation runs; in total 20 real-time runs were made.

Average annual returns from each of these runs are shown in columns 3 to 7 of Table 5.4. Standard deviations of annual costs are in corresponding positions in Table 5.5.

Real-time operating costs in Table 5.4 are estimates of the expected average annual costs for operating the model, as distinct from the absolute average annual costs given by the deterministic solutions. Provided that the linear policy functions used here are the appropriate way to express the real-time policy, the expected costs obtained here should be equivalent to thpse from a direct stochastic dynamic programming solution, were such a solution feasible.

There is a small amount of variation in expected costs in Table 5.4 for different real-time policies. In particular, the policy derived from data set 1 always gives greater costs than do policies from the other data sets. On the other hand, the standard deviation of costs with the policy from data set 1, is always less than the standard deviation of costs with other policies (Table 5.5). This particular policy might be described as more conservative than the others.

The variability in expected costs along the rows of Table 5.4 represents variability from using a linear policy estimated from sample data, rather than a linear policy based on population data. To reduce this variability, the standard length of 40 years used in estimating a sample policy must be increased.

5.5.2 Comparison of Deterministic Solution and Real-Time Operation

The expected costs in Table 5.4 are on average 28% greater than the corresponding absolute costs. We note here that this definite distinction between expected and absolute costs was not present in the example reported in Chapter 4.

The gap between expected and absolute costs might be narrowed by finding functional forms of (5-11) that fit the optimal data better than the linear forms (5-12). This possibility has already been touched on by trying quadratic functions, but no significant improvement in the explained variation of d_A and d_B was achieved. To the author, it seems unlikely that other functional forms will be improvements over the linear forms.

The deterministic solution for a four year period has been plotted with full lines in Figures 5.3 and 5.5. For comparison, the equivalent segment from a real-time operation run, that using the policy derived from data set 4, is shown by dashed lines in Figures 5.3 and 5.5. Annual totals for the two sets of thermal costs are given in Table 5.6.

Through years 10 and 11, the two solutions follow each other closely. At times the two storage curves depart from each other, but in many instances this happens in the opposite direction for each storage. Examples of this are between months 4 and 9 of year 10 and months 7 and 10 of year 11 in Figure 5.3. In these instances the total potential energy stored in the system by each solution is much closer than is apparent at first glance.

The real-time operation through the dry period of year 13 incurs a high cost relative to the deterministic solution (Figure 5.7). This is not a weakness of the real-time policy as such. Rather it is an illustration of the inability of real-time operating policies to allow for future dry spells to the same extent as the deterministic solution. Thus in year 12, one before the dry year, the thermal cost for real-time operation was $123,000; this is $86,000 less than that for the deterministic solution. However, through this year the deterministic solution built up substantial reserves, and the costs incurred through the following dry year were much less than those for the real-time operation (see Figure 5.7 and Table 5.6).

5.6 FLOW FORECASTING

An important advantage of the deterministic algorithm over other solution techniques is that flow forecast information can be readily utilized by incorporating it as "independent" variables in the policy functions (5-11). Value judgements on the worth of forecasts of a given reliability can be made by assessing real-time operating costs with and without forecasts. A balance exists between costs of obtaining more reliable forecasts and reductions achieved in operating costs (Young, 1966).

At the beginning of each month, two information sources can contribute to forecasts:

(1) To the extent that serial correlation exists between flows, the flow q_{n-1} in the (n-1)th month can be used to obtain a least squares estimate of the flow q_n in the nth month: thus assuming a covariance stationary model

for a single station, from (2-21) we have

$$\tilde{q}_n - \bar{Q}_j = \frac{S_j}{S_{j-1}} \cdot r(1) \cdot (q_{n-1} - \bar{Q}_{j-1})$$

where $\tilde{q}_n$ is an estimate of q_n, made at the beginning of the nth month. In the present case, incorporation of q_{n-1} terms into the policy functions did not improve estimates of d_A and d_B. This might have been expected a priori, since serial correlations for flows within the Waitaki were low (Table 3.6).

(2) Additional information X_n, known at the beginning of the nth month may be used to forecast flows one or more months ahead. Thus in general,

$$\tilde{q}_n = \tilde{q}\,(q_{n-1}, X_n)$$

In the Waitaki, the most likely source of the information X_n will be from snow surveys. Preliminary snow surveys have been undertaken in the area for the last five years. The data obtained in these surveys indicate the difficulties of snow surveying in New Zealand conditions. As given by Chinn (1970), these are;

(a) The uneven distribution of snowpack. This unevenness occurs because the snow catchments are exposed to strong prevailing winds which re-deposit snowpack, and because winter melting takes place on north facing slopes.

(b) The erratic build-up of snowpack in the winter. Rain and melting are likely to occur at any time and there is no particular date for maximum snow accumulation. The variability in runoff from rainfall can overshadow the variability in runoff due to snowmelt, making validation of forecasts difficult.

This contrasts with continential regions where snow runoff forecasting is carried out. Continential climates give an almost steady snowpack build-up to a maximum, followed by a steady depletion. This takes place over a defined season with limited winter melt.

At present in New Zealand it is not feasible to attempt to forecast snowmelt runoff on the basis of snow-survey information.

5.7 EXTENSIONS OF THE MODEL

This chapter has demonstrated the power and versatility of deterministic dynamic programming coupled with streamflow simulation. No other optimizing approach is currently capable of deriving an optimum real-time policy for a two storage model whilst taking account of the complex stochastic hydrology of this system.

A more realistic model of the system would allow usable storage at Benmore as well, giving a three state variable problem. The computational difficulties implied by this are described in Appendix I, but a solution should still be feasible if the program refinements suggested in section 5.4.1 are adopted.

A further step towards reality would be a model including other hydro plants and other storages outside the Waitaki. The hydrology of this problem should not present difficulties. Provided that lag-one, log-normal models are appropriate for monthly flow series, the multistream flow generating model described in Chapter 3 can be applied.

Difficulty would arise in solving the functional equation: a direct solution is no longer feasible.

Fukao and Nureki (1962), describe a successive approximation technique for finding an optimum policy for a power system which incorporated hydro and thermal generation. Each hydroplant was on a separate stream, and the objective in operation was to minimize thermal costs. The technique relies on approximation in policy space; this works in the following way:

1. Assume an operating policy for each storage.
2. For one storage derive an optimum policy, using the assumed policies for the remaining storages.
3. For the particular storage, replace the assumed policy with the optimum policy.
4. Repeat (2) and (3) for each storage.

5. Repeat (2), (3) and (4) until operating costs converge to a steady value.

For the case examined, convergence to a steady state was found to be quite rapid. This technique of successive approximations is worth investigating with a view to applying it to the system described herein. The weakness of this technique is that it gives no general assurance that the policies derived are associated with a global optimum, and not a local optimum.

5.7.1 Further Analyses

With the models described, the analyses undertaken can be extended in many ways, especially with respect to using alternative estimates for many of the system parameters. For instance;

(1) Only one set of energy loads was considered. Others are possible, differing both in their magnitude and in their month-to-month pattern.

(2) The dam at Pukaki has yet to be built. After finding an optimal policy for each of a range of storage sizes for this dam, a benefit-cost analysis could be undertaken to define an optimal dam size. In addition, alternative generating capacities for stations 2, 3, 4 and 5, which have yet to be built, could also be examined.

(3) Thermal costs were crudely approximated with a quadratic function. Intuitively, this shape is correct, but work is required to define the costs more clearly.

(4) Only one value of the interest rate was used: others are possible. However, this parameter should not be important. The problem dealt with here has been the allocation of water between months, and particularly through dry spells which at the most last for 12 to 18 months. This differs from the irrigation system examined in Chapter 4, where the problem was year-by-year water allocation, and the time value of money was relatively more important.

5.8 SUMMARY

This chapter has described the derivation of a long term storage operating policy for a model of a hydro-thermal power system.

Certain simplifications, necessary to set up the model, restrict the applicability of the derived policy to the actual operating problem.

Nevertheless, the approach used here, that of simulation - deterministic optimization - linear regression, holds promise for tackling more complex models that are closer to the real system.

An important advantage of the approach used is its ability to cope with a complex stochastic hydrology. Although lag-one Markov models were appropriate here, there is no difficulty in handling flows which require more complex models.

CHAPTER SIX

SUMMARY, DISCUSSION AND CONCLUSION

6.1 THESIS SUMMARY

Chapter One of this thesis outlined the process of planning for water resources development. It postulated that the formulation of a policy for real-time operation of storages is part of this planning process. Alternative optimizing techniques which have been used in water resource system design were briefly reviewed.

One way to take account of the stochastic nature of river flows in deriving a real-time policy was to use a three step process, coupling streamflow simulation and deterministic dynamic programming with linear regression. This process was illustrated in Figure 1.3.

Chapters Two and Three contain statistical analyses of monthly flow data from some New Zealand rivers.

Chapter Two describes the analysis and synthesis of single data series. It was shown that the monthly flow series can be represented by lag-one lognormal Markov models. Important statistical parameters in this analysis and synthesis were the monthly means, standard deviations and skewness coefficients, and the lag-one serial correlation coefficients.

When considering flow series for several streams within a region, between-station correlations represent parameters additional to those for the single station models. Chapter Three described and tested a multi-stream model which operated in terms of the at-station parameters and the between-station correlations.

Chapters Four and Five gave two examples of the application of the real-time policy derivation technique.

An introductory example in Chapter Four involved year-by-year decision-making for releases from an irrigation dam. Operation with the derived real-time policies was compared with deterministic solutions and with operation under the currently used policy. Differences of less than 1% were found between expected maximum returns using the derived real-time policy and absolute maximum returns from the deterministic solution. Further, in terms of annual returns, there was little difference between the derived policy and that which is currently used.

A hydro-thermal power system was examined in Chapter Five. For this system the multistream model developed in Chapters Two and Three was used to simulate interrelated sequences of flows. The real system was approximated by a two storage model, for which real-time policies were derived; the objective in operation being taken as the minimization of thermal generating costs.

In comparing the real-time policies with deterministic solutions, it was noted that expected minimum costs were substantially greater than absolute minimum costs. This cost difference, which can be described as the cost of uncertainty in flows, could be reduced by forecasting flows one or more months ahead. In future, it is likely that snow survey data will provide information for forecasting. At present, however, difficulties in establishing snow survey methods make runoff forecasting impossible.

6.2 DISCUSSION

Some of the points mentioned in summary warrant further discussion.

Time series analyses of monthly flow data from seven New Zealand rivers are reported. Lag-one log-normal Markov models fit these data. These models can be used to generate synthetic data. Basic parameters of an historic record which are shown to be reproduced in the synthetic data are the monthly means and standard deviations, the skewness coefficient and the lag-one serial correlation coefficient.

Difficulty was encountered in testing the significance of the differences between historic and synthetic means and standard deviations. This point, covered in detail in Section 3.4.4, arose because the significance test first used required estimates of means and standard deviations to be normally distributed about population values. But this requirement is only approximately met when log-normal distributions are used and thus the confidence limits established for parameter estimates are only approximate. With a modified test, good agreement between historic and synthetic means and standard deviations was shown in Figures 2.11, 2.12 and 2.13.

In three cases some evidence was obtained supporting the use of lag-two instead of lag-one models. However, the evidence was slight, and was insufficient to justify the additional complexity. In all but two cases, covariance stationary models appeared to be reasonable.

A multi-stream generating method due to Matalas (1967b) is described in detail and results from applying it to four flow stations within the Waitaki River basin are given. Not only are the at-station parameters preserved by this procedure, but also the interstation correlations are accounted for.

Previous results from applying this method (Young and Pisano, 1968) are not satisfactory on the grounds that there is excessive variation between historic and synthetic parameters. This variation appears to be due to a failure to transpose a matrix. The results reported herein are an improvement.

The latter part of this thesis illustrates the use of flow simulation coupled with deterministic optimization in deriving real-time operating policies. To do this, two water storage systems are examined, and policies are obtained for their operation. In discussion it is interesting and instructive to ask questions such as "What are the data required in this type of study?" "In what directions should efforts be extended in collecting data?" "How do the models used here approximate the real situations?"

"What are the model limitations?"

For the optimizing studies described herein, the data requirements are quite specific. These are hydrologic and economic. In the present context hydrologic data means flow, rainfall, evaporation and seepage figures, and storage-elevation curves. Economic data requirements are implied by the design criteria listed in Section 1.2. When structural sizes and levels of output have been fixed, as was the case in these studies, and the problem is to define a policy for real-time operation, the list of economic data is substantially reduced to (a) an interest rate (or a range of possible values) and (b) a short run net benefit or benefit loss function. In the past relatively little attention has been given to this type of function which puts values on the losses incurred from failures to meet planned or target outputs.

Data availability was one of the main criteria in choosing the particular examples studied, and definitions of net benefit or benefit loss functions were readily obtained. For the irrigation dam, this function followed directly from the contracted agreement of charging for quotas and reducing the charge when water deliveries are short. In the hydro system it was given by costing alternative thermal energy sources.

It is important to emphasize that these studies were operational, and that for the purposes of deriving operating policies the absolute values of costs or benefits are not important, provided the functional form of the benefit or benefit loss function is known.

This is not the case in design studies where both structural sizes and output levels are to be optimized; here benefits and costs must be clearly defined in absolute terms. In New Zealand work is needed to define the benefits obtained from different water uses and water products and to quantitatively evaluate the cost of failures to meet planned or target output levels. Until this is done, optimization studies which involve inputs of capital cannot be carried through to completion.

Nevertheless, the approach examined here can and should be applied to other storage systems where one, two or three storages are involved and where real-time operating policies are required.

To evaluate the usefulness of the models described here in actual design and operation studies, judgement and experience are required. The following comments are pertinent however.

The irrigation dam examined was a very simple system and its main features have been incorporated in the model. The policy obtained compared closely with that which is currently used. If anything it was a slight improvement. However, because the currently-used policy greatly reduced the risk of failure of anticipated supplies, it is to be preferred.

In certain respects the hydro-thermal system examined here is only a simplified model of a very complex interconnected generating system. Its major limitation is the restriction to two storages. It is this feature that restricts the applicability of the derived policy to operation of the real system. Although conceptually straightforward, incorporation of the third storage at Benmore would lead to computational restrictions.

Other limitations are the assumptions of known monthly energy loads and of full availability of generating plant.

Nevertheless, the basis of the model is realistic and important advantages of the approach are illustrated. Most notable is the complete account taken of an interrelated stochastic hydrology; no other optimizing approach is capable of this. Although lag-one Markov schemes were appropriate here, no restriction is placed on the use of other more complex schemes for simulating flow data. Other features that are readily handled are losses from storages, the time value of money and variable head power stations.

Although not undertaken in the study, variable allocations through the year of empty space for flood control storage could be allowed for by adjusting the active storage volumes. In the Waitaki, floods are most likely to occur in the summer and at this time of year space for flood control could be made available by reducing the volumes allocated for active storage. Through the winter, when floods are less likely, flood control storage space could be reduced and the volumes allocated for active storage increased. Such operations should only have minor effects on long-term operation however, since generally the storages fill in summer and empty through the winter, and the only important parameters will be the active storage volumes made available in the summer months.

In that it applies dynamic programming to the seasonal operation of a two-storage instead of a one-storage hydro-thermal power system, this study is an extension of the work of Little (1955). Also, it extends the work of Young (1966) in applying simulation coupled with deterministic dynamic programming and multiple regression to a realistic stochastic operation problem.

More complex models involving storages outside the Waitaki basin would give formulations that cannot be solved directly. For these situations, the successive approximation techniques described by Fukao and Nureki (1962) are worth investigation.

6.3 CONCLUSIONS

(1) Lag-one, log-normal Markov models gave good fit for the monthly flow series from seven recording stations. In all but two instances, covariance stationary models were appropriate. In three instances there could be some justification for lag-two models.

(2) Streamflow synthesis should be preceded by analysis to determine an appropriate model.

(3) The three step deterministic optimization approach applied in real-time policy derivation is a very versatile technique. In particular;

(a) it is not restricted to a linear objective function, but as used here, a smoothly and monotonically increasing (or decreasing) objective function was implied by finding a functional form for the policy;

(b) the time value of money, which influences the real-time policy, is easily accounted for;

(c) seepage and evaporation from storages are readily coped with;

(d) variable head power stations present no difficulty;

(e) flow forecasts can be incorporated into the policy function; (In addition, their worth can be quantitatively assessed.)

(f) an absolute optimum policy for deterministic inflows can be obtained and used as a yardstick in examining real-time policies;

(g) the derived real-time policy need not be restricted to linear functional forms; (However, in the examples herein, linear forms were found to be appropriate)

(h) constraints restricting the range of allowable storage and release values present no difficulties. They have the advantage of reducing the computational effort required.

(i) stochastic interrelated hydrologies present no difficulties. Interrelated flow series requiring models more complex than Markov lag-one could be coped with - provided an appropriate multivariate model could be formulated.

(4) The major limitation of the three step deterministic optimization approach is computational; direct solutions are only feasible for systems which incorporate one, two or perhaps three storages.

(5) Successive approximation techniques may be able to reduce this limitation. These are worth further investigation.

REFERENCES

AITCHISON, J. and BROWN, J.A.C. (1957) The Lognormal Distribution. Cambridge Univ. Press.

ASKEW, A.J., YEH, W.W-G. and HALL, W.A. (1970) "Streamflow Generating Techniques: A comparison of their abilities to simulate critical periods of drought." Univ. of California Water Resources Centre Contribution No.131.

BARNES, F.B. (1954) "Storage Required for a City Water Supply." J.I.E.Aust. V. 26 (9), pp 198-202.

BEARD, L.R. (1965) "Use of Interrelated Records to Simulate Streamflow." Proc. ASCE, V.91, HY5, pp 13-22.

BRITTAN, M.R. (1961) "Probability Analysis Applied to the Development of Synthetic Hydrology for the Colorado River." Part IV of "Past and Probably Future Variations in Stream Flow in the Upper Colorado River." Bureau of Economic Research, Univ. of Colorado.

BROOKS, C.E.P. and CARRUTHERS, N.C. (1953) Handbook of Statistical Methods in Meteorology. H.M.S.O. London.

BURAS, N. (1965) "A Three Dimensional Optimization Problem in Water Resources Engineering." Operations Research Quarterly. V 16 (4).

_________ (1966) "Dynamic Programming in Water Resources Development." Advances in Hydroscience, V 3, pp 367-412.

_________ and HERMAN, T. (1968) "A Review of Some Applications of Mathematical Programming in Water Resources Engineering." (progress report no. 2) Technion - Israel Institute of Technology.

BURTON, J.R. (1964) "Hydro-Economic Planning for Australia's Water Resources." pp 438-449 in Water Resources Use and Management. Melbourne Univ. Press.

CHINN, T.J. (1970) Personal communication.

CHOW, V.T. (1964) (editor) Handbook of Applied Hydrology, McGraw-Hill.

DORFMAN, R. (1965) "Formal Models in the Design of Water Resource Systems." J. Water Resources Res., V 1(3), pp 329-336.

ECKSTEIN, O. (1958) Water Resources Development, The Economics of Project Evaluation. Harvard Univ. Press.

FIERING, M.B. (1964) "Multivariate Techniques for Synthetic Hydrology." Proc. ASCE, V 90, HY5, pp 43-60.

__________ (1967) Streamflow Synthesis. Harvard Univ. Press

__________ (1968) "Schemes for Handling Inconsistent Matrices." J.Water Resources Res., V 4(2), pp 291-297.

FUKAO, T. and NUREKI, R. (1962) "Successive Approximations Methods of Dynamic Programming in Reservoir Control Problems." Bulletin of the Electrotechnical Laboratory, Japan, V 26(3) pp 18-42.

HALL, W.A. (1964) "Optimum Design of a Multiple-Purpose Reservoir" Proc.ASCE, V 90, HY4, pp 141-149.

__________ and BURAS, N. (1961) "The Dynamic Programming Approach to Water-Resources Development." J. Geophys. Res. V 66(2), pp 517-520.

__________, BUTCHER, W.S. and ESOGBUE, A., (1968) "Optimization of the Operation of a Multi-Purpose Reservoir with Dynamic Programming." J. Water Resources Res. V 4(3), pp 471-477.

__________ and HOWELL, D.T. (1963) "The Optimization of Single-Purpose Reservoir Design with the Application of Dynamic Programming to Synthetic Hydrology Samples" J. Hydrology, V 1, pp 355-363.

__________ and ROEFS, T.G. (1966) "Hydropower Project Output Maximization" Proc. ASCE, V 92, PO1, pp 67-79.

HARMS, A.A. and CAMPBELL, T.H. (1967) "An Extension to the Thomas-Fiering Model for the Sequential Generation of Streamflow." J.Water Resources Res. V 3(3), pp 653-661.

HAZEN, A. (1914) "Storage to be Provided in Impounding Reservoirs for Municipal Water Supply." Trans. ASCE, Vol. LXXVII, pp 1539-1640.

HUFSCHMIDT, M.M. (1965) "Field Level Planning of Water Resource Systems." J. Water Resources Res. V 1(2), pp 147-164.

______________ and FIERING, M.B. (1966) Simulation Techniques for Design of Water Resource Systems. Harvard Univ. Press.

JENNINGS, M.E. (1969) Discussion with Young and Pisano (1968), Proc. ASCE, V 95 HY4, pp 1468-1469.

JENSEN, R.C. (1968) "Economic Evaluation of Water Resources Development." Lincoln Papers in Water Resources, No 2, pp 22-40, Lincoln College, Canterbury, N.Z.

JULIAN, P.R. (1961) "A Study of the Statistical Predictability of Runoff in the Upper Colorado River Basin," part II of "Past and Probable Future Variations in Stream Flow in the Upper Colorado River." Bureau of Economic Research, Univ. of Colorado.

KENDALL, M.G. and STUART, A. (1968) The Advanced Theory of Statistics, Vol.3, Design and Analysis, and Time Series, Charles Griffin, London.

LEWIS, D.J. and SHOEMAKER, L.A. (1962) "Hydro System Power Analysis by Digital Computer" Proc. ASCE, V 88, HY3, pp 113-130.

LITTLE, J.D.C. (1955) "The Use of Storage Water in a Hydroelectric System" J. Operations Research Society of America, V 3(2), pp 187-197.

McKEAN, R.N. (1958) Efficiency in Government through Systems Analysis, Wiley, N.Y.

MAAS, A.A. et al (1962) Design of Water Resource Systems. Harvard Univ.Press.

MANDELBROT, B.B. and WALLIS, J.R. (1968) "Noah, Joseph, and Operational Hydrology." J. Water Resources Res., V 4(5), pp 909-918.

MATALAS, N.C. (1967a) "Time Series Analysis" J. Water Resources Res, V 3(3) pp 817-829.

____________ (1967b) "Mathematical Assessment of Synthetic Hydrology" J. Water Resources Res., V 3(4), pp 931-945.

_____________ and BENSON, M.A. (1968) "Note on the Standard Error of the Coefficient of Skewness." J. Water Resources Res., V 4(1), pp 204-205.

MOREAU, D.H. (1968) "A Comparative Examination of Alternative Definitions of Serial Correlation in Streamflow Synthesis." American Water Resources Assn. Proc. of the National Symposium on the Analysis of Water Resource Systems. pp 298-308.

MORRICE, H.A.W. and ALLAN, W.N. (1959) "Planning for the Ultimate Hydraulic Development of the Nile Valley." Proc. Inst. Civil Engrs, V 14, pp 101-156.

QUIMPO, R.G. (1967) "Stochastic Model of Daily River Flow Sequences." Hydrology Paper No.18, Colorado State Univ. Fort Collins, Colo.

RODRIQUEZ - ITURBE, I. (1969) "Estimation of Statistical Parameters for Annual River Flows." J. Water Resources Res., V 5(6), pp 1418-1421.

ROEFS, T.G. (1968) Reservoir Management: the State of the Art. IBM Washington Scientific Research Centre, 320-3508.

ROESNER, L.A. and YEVDJEVICH, V.M. (1966) "Mathematical Models for Time Series of Monthly Precipitation and Monthly Runoff Series." Hydrology Paper No 15, Colorado State Univ. Fort Collins, Colo.

SCHWEIG, Z. and COLE, J.A. (1968) "Optimum Control of Linked Reservoirs." J. Water Resources Res., V 4(3), pp 479-497.

SUDLER, C.E. (1927) "Storage Required for the Regulation of Streamflow." Trans. ASCE, V 91, pp 622-660.

THOMAS, H.A. and FIERING, M.B. (1962) "Mathematical Synthesis of Streamflow Sequences for the Analysis of River Basins by Simulation." Ch. 12 in Maas et al (1962).

YEVDJEVICH, V.M. (1964) "Fluctuations of Wet and Dry Years, Part II, Analysis by Serial Correlation." Hydrology Paper No.4, Colorado State Univ. Fort Collins, Colo.

YOUNG, G.K. (1966) Techniques of Finding Reservoir Operating Rules. Ph.D. thesis, Harvard Univ.

__________ and PISANO, W.C. (1968) "Operational Hydrology using Residuals," Proc. ASCE, V 94, HY4, pp 909-923.

APPENDIX I

COMPUTER PROGRAMS AND COMPUTING DETAILS

INTRODUCTION

To obtain the results described in this thesis, a considerable effort was required in all phases of writing, checking, "de-bugging", and executing several large computer programs. The intention of this appendix is not to give detailed descriptions of these programs, but rather to briefly describe the computing facilities used, and to summarize the main steps of the programs and to discuss programming features that the author found relevant to the solution of his deterministic dynamic programming problems.

COMPUTING FACILITIES

Two different computer installations, the IBM 1130 at Lincoln College, and the IBM 360 model 44 at Canterbury University, were used in this study.

The IBM 1130 has a capacity of 8192 addressible 16 bit words and is fitted for card input and card and printer output. It has one disk storage device which can store 512000-16 bit words. In the 1130 system real (single precision) numbers take 2 words of storage - thus the core and disk have respective capacities equivalent to 4096 and 256000 real numbers. Allowance for permanently stored system programs reduces the space available to a user to about 3000 and 200000 real numbers respectively.

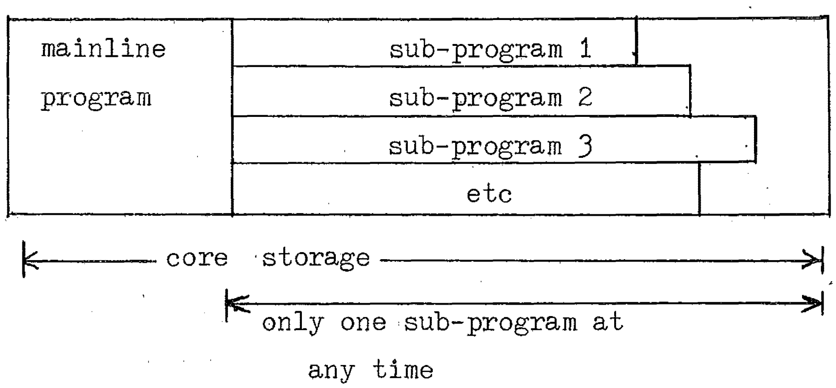
The IBM 360/44 is a much more powerful installation. It has an available core capacity of about 25000 single precision (4 byte) real numbers, back-up disk storage facilities of about 300,000 single precision numbers and greatly improved speeds of compilation, processing and output.

Both computers were equipped with FORTRAN IV compilers, the version for the 360/44 being more general than that for the 1130. Programs written for the 1130 could be transferred to the 360/44, but not necessarily vice-versa. All the programming undertaken in this study used FORTRAN IV.

Both installations had allowance for "phase overlaying" of programs. With this facility, computers with limited core but extensive backing storage can handle quite long programs. This is achieved in the following way.

The long program is segmented into a number of sub-programs and a co-ordinating mainline program, and these are all stored on disk. During execution, control starts in the mainline; as different segments of the program are required to be executed, the appropriate sub-program is called into core storage from disk and control is passed to it. When finished, control passes back to the mainline from which another sub-program may be called. The important point is that when a sub-program is loaded into core from disk, it enters the space occupied by the previous sub-program. Thus at any time only one sub-program is in core storage, and sub-programs overlay one another. If the sub-programs all require approximately the same quantity of storage, considerable economies are achieved, since storage needs only be set aside for one, instead of several sub-programs.

The process is illustrated as follows:



In the 360 system this scheme is termed a "root phase overlay", while a similar process in the 1130 system refers to overlaid subroutines as LOCAL (load-on-call) subroutines. The programs written have made extensive use of this feature.

Disk storage is also important for data storage; in solving dynamic programming problems, tables of optimal releases for each stage can be stored on disk and recalled later.

DESCRIPTION OF PROGRAMS

In the course of the study a number of computer programs were written. In the analysis and synthesis of flow data for single stations several short programs were used. These were straightforward statistical programs and do not merit any comment.

The three major programs written were for;

(a) the multistream flow simulation

(b) the operation of an irrigation storage, and

(c) the operation of the two-storage hydro-thermal power system.

The important steps in these programs are briefly described.

(a) Multistream Program

The multistream analysis and synthesis program mentioned in Chapter Three was written for the 1130 and was subsequently implemented with the 360/44. The major steps in this program were:

(1) Read historic flow data.

(2) Analyse historic data, printing $\bar{Q}_j$, S_j, C_{sj}, C_s, r_j, $r(1)$ and cross-correlations.

(3) For each station assume

(a) log-normal distribution

(b) covariance stationarity

(c) lag-one Markov process

(4) Check consistency of correlation matrix.

(5) From (3-5) and (3-6) calculate matrices A and B.

(6) With (3-1) generate synthetic Y_t.

(7) With the transform (3-2) introduce periodicity and take antilogs to form synthetic flows.

(8) Analyse synthetic data, and compare synthetic parameters with historic parameters.

The major restriction of this program is that the assumptions in step (3) must be tenable for all the stations. In the event of a large number of stations being considered, computational restrictions might also apply.

(b) Program for Operation of Irrigation Storage

The program used was a modified version of SNAKE GULLY, a storage operation program developed by J.R. Burton. SNAKE GULLY considered month-by-month operation and was written for an IBM 700 series computer, but this program worked on a year-by-year basis and used the much smaller IBM 1130.

Steps in the program were (using the notation of Chapter Four):

(1) Read input data-storage capacity, step sizes, interest rate, flow parameters, length of analysis required, etc.

(2) Generate synthetic inflows and print synthetic parameters.

(3) Deterministic optimization.

(a) Start at last period, $n = 1$. Find $f_1(s)$ and corresponding d for all s, $0 \leqslant s \leqslant S_M$. Store the d values on disk.

(b) $n = n + 1$. From (4-4) find $f_n(s)$ and corresponding d for all s, $0 \leqslant s \leqslant S_M$. Store d values on disk. Repeat this step until $n = N$.

(c) Carry out storage behaviour analysis (forward mass balance). Start at first time interval (beginning of real-time), set $n = N$. Assume initial s value (in this case $s = \frac{1}{3} S_M$), retrieve optimal release table from disk and select d value appropriate to initial volume s. Hence compute s', the end-of-period volume. Repeat this step for n = N-1, N-2 ... 2, 1, printing s, d, q and annual returns.

(4) Use multiple regression to determine real-time policy functions of the type (4-5).

Separate sub-programs were used for generating synthetic flows (step 2), finding functional tables (steps 3a, 3b) and carrying ~~out~~ storage behaviour analyses (step 3c). Real-time operation runs were carried out with a modified version of the storage behaviour analysis sub-program.

These programs were implemented with the IBM 1130. The execution time for obtaining optimal releases and determining a year-by-year policy for a 48 year sequence of simulated data was of the order of 20 minutes, the exact time depending on the detail required in the output.

(c) <u>Program for Two-Storage Hydro-Thermal Power System</u>

This program is a "one-off" job - it applies only to the Waitaki system. The basic form of this program is similar to the single storage SNAKE GULLY program, but is beyond the capacity of the 1130 system and was written with the features of the 360/44 system in mind.

Steps in the program were;

(1) Implement multistream program to generate 41 years of synthetic data.

(2) Deterministic optimization.

(a) Start at last time interval, set n = 1, and find $f_1(s_A, s_B)$ and corresponding d_A and d_B for all feasible s_A and s_B. Store d_A and d_B tables on disk.

(b) n = n+1. From (5-6) find $f_n\ (s_A,\ s_B)$ and store corresponding d_A and d_B on disk. Repeat this step until n = N. For selected n values, print the tables of $f_n\ (s_A,\ s_B)$ and d_A and d_B.

(c) Carry out storage behaviour analysis (or forward mass balance). Print storage volumes, releases, inflows, losses, quantities of hydro and thermal energy generated and the thermal costs incurred.

(3) Use multiple regression to determine real-time policy functions.

The complete program comprised a coordinating mainline program and four overlaid sub-programs. The sub-division of the program is summarized in the following table:

Mainline Program ("Root Phase")	Sub-Programs (Overlaid Phases)		
	Name	Steps (as listed above)	Task
DYNAPRO	DATAGEN	1	data generation
	DETERM	2a, 2b	determination and storage of d_A and d_B tables
	BEHAVE	2c	storage-behaviour analysis, - finds optimal s_A, s_B, d_A, d_B
	POLICY	3	multiple regression to find policy functions of the type (5-12)

In testing real-time policies, a slightly modified version of the storage-behaviour program BEHAVE was used in conjunction with the multistream program DATAGEN.

COMPUTATIONAL REQUIREMENTS OF DYNAMIC PROGRAMMING

Many water resource planning problems can readily be formulated as dynamic programming problems; the difficulty is in obtaining numerical answers from the formulations, - that is, in writing a computer program that will fit into an available computer and that will not use excessive quantities of computer time. This difficulty is computational, not conceptual.

The space and time requirements of a program for solving a dynamic programming problem are two facets of the computational problem, and both must be considered in assessing the feasibility of a proposed computer program. In both these features Young (1966) showed that his deterministic (Monte Carlo) dynamic programming algorithm compared favourably with the stochastic algorithm.

Only one, two or possibly three state variables can be handled in direct solutions of deterministic problems. The reason for this is that both the space and time requirements of a computer increase geometrically with the number of state variables. The increase in space requirements can be demonstrated as follows:

Consider a general problem of N stages involving m state variables. Set the subscripts A, B, C, ... refer to the 1st, 2nd, 3rd, ... state variables, and let each state variable have K feasible values. (This discussion proceeds in the context of the storage operation problems examined, and the number of decision variables equals the number of state variables.)

Thus K^m combinations of values of state variables are possible, and to each combination there corresponds a value of the optimum return function $f_n(s_A, s_B \ldots)$ as well as the decisions d_A, d_B ... . At any stage, core storage must be provided for these quantities as well as f_{n+1} $(s_A, s_B, \ldots)$, and hence $(2 + m)\ K^m$ storage locations are required in core. Further, at every stage, all the d_A, d_B ... must be stored in backing storage for recall later in the storage behaviour analysis. Thus for an N stage process, the backing storage requirement is $N \cdot m \cdot K^m$.

For different values of m, these storage requirements are tabulated as follows for magnitudes of N and K that are the same order as those for the hydro-thermal problem.

STORAGE REQUIREMENT FOR DYNAMIC PROGRAMMING

PROBLEMS WITH N = 500 STAGES

K (no. of feasible values of state variable)	m (no. of state variables)	$(2 + m) K^m$ (core storage requirement)	$N m K^m$ (backing storage requirement)
20	1	60	10,000
	2	1,600	400,000
	3	40,000	12,000,000
	4	960,000	320,000,000
10	1	30	5,000
	2	400	100,000
	3	5,000	1,500,000
	4	60,000	20,000,000

This analysis illustrates why **direct solutions** with medium sized computers such as the 360/44 are limited in terms of storage capacity to problems involving only two or three state variables. Even with a relatively coarse grid of 10 feasible values for each state variable, and a drastic reduction in the number of stages to say 50, this computer could not handle a four state variable problem since both core and backing storage requirements would still exceed the capacity.

This analysis is extremely simplified, it does not take account of storage requirements for other variables such as release and storage steps, losses etc. Some of these can either be stored as arrays or recalculated each time they are needed, but others are best stored permanently. Further, it should be pointed out that not all the core storage available to a user can be used to store data, a portion is required for the actual program instructions.

Once a program has been written that will fit the space limitations of a computer, the question must be asked "Will the program run within a reasonable time?" There is no general answer to this question. One must decide what is a "reasonable" time and this is dependent on the economics of individual computer installations.

For the two state variable hydro-thermal program, execution time requirements given in Section 5.4.1 were substantial. Before a third state variable (for active storage at Benmore) could be introduced, coarser grids to reduce the number of feasible values of state variables would be required to reduce the storage requirements. This would also help to speed up the program and further improvements in speed could be obtained by rewriting sections of the program in ASSEMBLER language. As was mentioned in Chapter Five, the grid sizes used in the two state variable program gave an execution time that was just acceptable.

APPENDIX II

TABULATION OF LINEAR POLICIES

DATA SET (1)

Storage A $d_A = a_1 s_A + a_2 s_B + a_o$ | Storage B $d_B = b_1 s_A + b_2 s_B + b_o$

Month	a_1	a_2	a_o	R	SE	F	mean release	Month	b_1	b_2	b_o	R	SE	F	mean release
1	.15	-.05	70	.26	50	1.4	83	1	-.06	.20	108	.23	74	1.1	114
2	.40	-.12	38	.62	39	11.3*	65	2	.05	.12	137	.42	49	4.0*	205
3	.42	-.07	32	.43	38	4.1*	94	3	.06	.08	195	.33	43	2.3	256
4	.51	-.01	-45	.41	50	3.7*	69	4	-.01	.06	281	.26	35	1.4	312
5	.27	-.08	60	.44	42	4.5*	85	5	-.15	.19	258	.65	45	13.5*	314
6	.35	-.05	12	.36	51	2.7	70	6	-.17	.20	268	.70	32	17.6*	302
7	.01	-.15	118	.44	42	4.3*	90	7	.00	.23	191	.70	31	17.9*	236
8	.25	-.04	62	.51	30	6.5*	97	8	-.04	.23	178	.61	31	11.0*	196
9	.19	-.20	93	.67	24	15.4*	103	9	-.16	.23	175	.63	28	12.5*	170
10	.32	-.24	53	.45	34	4.7*	60	10	-.26	.22	140	.35	39	2.6	136
11	.05	-.20	82	.29	45	1.8	71	11	.02	-.08	127	.10	49	.2	123
12	.11	.08	35	.30	49	1.8	62	12	-.19	.02	152	.27	59	1.4	132

* Significant at 95% level

DATA SET (2)

Storage A $d_A = a_1 s_A + a_2 s_B + a_o$								Storage B $d_B = b_1 s_A + b_2 s_B + b_o$							
Month	a_1	a_2	a_o	R	SE	F	mean release	Month	b_1	b_2	b_o	R	SE	F	mean release
1	.22	.00	29	.33	52	2.3	68	1	-.10	.06	152	.20	69	.8	156
2	.14	.04	58	.17	48	.6	80	2	-.02	.06	191	.23	49	1.0	214
3	.38	-.09	57	.45	44	4.7*	96	3	.01	.08	231	.41	36	3.7*	281
4	.17	-.05	81	.21	47	.9	93	4	.22	.22	107	.67	45	14.9*	296
5	.39	-.05	23	.37	45	3.0	88	5	-.09	.13	278	.44	46	4.4*	336
6	.38	-.11	56	.56	39	8.5*	90	6	-.25	.21	279	.70	39	17.4*	319
7	.29	-.12	68	.47	37	5.3*	87	7	-.14	.18	242	.57	35	9.0*	270
8	.03	-.09	100	.24	42	1.1	91	8	.24	.15	162	.59	37	10.1*	221
9	.28	-.15	71	.60	35	10.4*	87	9	-.04	.26	167	.62	35	11.3*	183
10	.15	-.05	75	.26	36	1.4	83	10	-.03	-.05	155	.13	36	.3	150
11	.19	-.03	67	.27	42	1.4	77	11	-.31	-.14	176	.42	56	4.0*	145
12	.09	.01	62	.13	55	.3	73	12	-.17	-.03	171	.22	62	.9	149

* Significant at 95% level

DATA SET (3)

	Storage A	$d_A = a_1 s_A + a_2 s_B + a_o$							Storage B	$d_B = b_1 s_A + b_2 s_B + b_o$						
Month	a_1	a_2	a_o	R	SE	F	mean release	Month	b_1	b_2	b_o	R	SE	F	mean release
1	.19	-.03	55	.30	50	1.8	74	1	-.39	.02	209	.60	41	10.5*	158
2	.17	-.02	66	.24	47	1.2	90	2	-.16	.08	189	.32	51	2.1	194
3	.25	-.09	80	.49	38	5.8*	82	3	.03	.11	196	.59	31	9.7*	270
4	.27	-.02	29	.28	47	1.5	82	4	-.06	.09	263	.33	43	2.3	306
5	-.10	-.05	139	.22	45	.9	91	5	.21	.07	241	.34	44	2.5	328
6	.30	-.23	50	.43	43	4.1*	89	6	-.08	.13	266	.56	37	8.3*	302
7	.05	.02	96	.15	29	.4	110	7	.12	.13	186	.61	32	11.2*	242
8	.46	-.10	61	.74	30	22.3*	100	8	-.02	.23	181	.67	41	15.2*	219
9	.29	-.18	69	.62	31	11.4*	70	9	.42	.27	125	.81	28	36.1*	184
10	.46	-.23	58	.60	36	10.1*	70	10	-.01	.24	129	.41	44	3.8*	142
11	.11	.07	75	.24	43	1.1	89	11	-.37	.16	167	.58	41	9.2*	127
12	.12	-.04	67	.15	55	.4	71	12	-.12	-.04	178	.22	51	1.0	159

* Significant at 95% level

DATA SET (4)

Storage A $d_A = a_1 s_A + a_2 s_B + a_o$								Storage B $d_B = b_1 s_A + b_2 s_B + b_o$							
Month	a_1	a_2	a_o	R	SE	F	mean release	Month	b_1	b_2	b_o	R	SE	F	mean release
1	.04	-.07	97	.28	51	1.5	82	1	-.10	.01	154	.17	57	.5	141
2	.28	-.06	40	.44	48	4.5*	66	2	-.39	.08	237	.44	65	4.4*	196
3	.75	-.09	-41	.65	43	13.7*	79	3	-.28	.53	306	.33	40	2.2	274
4	.51	-.05	-16	.34	50	2.4	82	4	-.43	.17	314	.59	41	9.8*	309
5	.21	-.04	48	.23	48	1.0	77	5	-.23	.11	323	.49	39	5.7*	326
6	.17	-.05	65	.26	48	1.4	85	6	-.08	.11	277	.45	37	4.6*	300
7	.35	-.18	57	.69	41	16.7*	82	7	-.03	.21	203	.78	29	29.0*	248
8	.06	-.15	109	.45	37	4.8*	99	8	.09	.22	161	.59	40	10.0*	204
9	.19	-.15	84	.53	32	7.3*	92	9	.09	.15	162	.43	35	4.3*	183
10	.12	.05	64	.25	38	1.3	74	10	-.08	-.11	152	.36	32	2.7	142
11	.12	-.01	67	.22	41	.9	75	11	-.20	.05	146	.26	60	1.3	134
12	-.08	.01	80	.13	52	.3	72	12	-.02	-.07	164	.17	65	.5	150

*Significant at 95% level

DATA SET (5)

Storage A $d_A = a_1 s_A + a_2 s_B + a_o$								Storage B $d_B = b_1 s_A + b_2 s_B + b_o$							
Month	a_1	a_2	a_o	R	SE	F	mean release	Month	b_1	b_2	b_o	R	SE	F	mean release
1	-.11	.03	77	.17	56	.6	67	1	-.21	-.08	193	.38	59	3.1	138
2	.17	-.06	81	.33	50	2.3	88	2	-.15	.09	177	.37	50	2.0	184
3	.24	-.08	78	.43	33	4.3*	85	3	-.30	.08	293	.44	35	4.5*	273
4	.11	-.01	83	.11	43	.2	105	4	.06	.10	217	.36	49	2.7	297
5	.45	.02	-42	.47	44	5.3*	69	5	-.11	.11	293	.39	46	3.3*	334
6	.16	.04	46	.47	45	5.3*	87	6	-.06	.15	254	.61	36	10.8*	304
7	.06	-.07	109	.27	44	1.5	101	7	.13	.17	177	.66	32	14.1*	250
8	.28	-.16	71	.61	38	10.9*	85	8	.04	.27	165	.77	29	26.3*	217
9	.39	-.22	71	.70	32	17.9*	91	9	.04	.40	154	.76	33	26.0*	195
10	.44	-.03	54	.61	30	11.2*	76	10	-.13	.12	159	.24	34	1.1	157
11	-.18	-.02	73	.23	44	1.1	62	11	-.08	-.11	160	.17	49	.6	151
12	.03	-.00	57	.05	54	.1	60	12	-.19	-.12	187	.31	61	2.0	156

* Significant at 95% level